



GRADUATE SCHOOL
ISTANBUL MEDENIYET UNIVERSITY
DEPARTMENT OF PRIVATE LAW

A Challenge to Copyright: Text and Data Mining

Master's Thesis

Esma Muheyne DOĞAN

May 2024



GRADUATE SCHOOL
ISTANBUL MEDENIYET UNIVERSITY
DEPARTMENT OF PRIVATE LAW

A Challenge to Copyright: Text and Data Mining

Master's Thesis

Esma Muheyne DOĞAN

Supervisor

Assoc.Prof.Dr. Cahit SULUK

May 2024

THESIS JURY APPROVAL

This master's thesis titled "A Challenge to Copyright: Text and Data Mining" written by Esma Muheyne DOĞAN at the Department of Private Law was accepted by our jury.

JURY MEMBERS

SIGNATURE

Supervisor:

Assoc.Prof.Dr. Cahit SULUK
Istanbul Medeniyet University, Faculty of Law

Other Jury Members:

Asst.Prof.Dr. Tuğba GÜLEŞ
Istanbul 29 Mayıs University, Faculty of Law

Asst.Prof.Dr.Abdurrahman Hamza TÜZGEN
Istanbul Medeniyet University, Faculty of Law

Date of Defense: 28.05.2024

STATEMENTS

Style and Reference Manual Statement

Having reviewed this thesis written under my supervision, I confirm that it has been written in accordance with Chicago Manual of Style and used its footnote reference format consistently throughout the entire text.

Signature

Assoc.Prof.Dr. Cahit SULUK

Declaration of Originality

I hereby declare that all information in this dissertation has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conducts, I have fully cited and referenced all material and results that are not original to this work.

Signature

Esma Muheyne DOĞAN

GENİŞ ÖZET

Telif Hukukunda Metin ve Veri Madenciliği

DOĞAN, Esmâ Muheyne

Yüksek Lisans Tezi, Özel Hukuk Anabilim Dalı

Danışman: Doç. Dr. Cahit SULUK

Mayıs 2024

Yapay zekâ, makine öğrenimi ve dolayısıyla metin ve veri madenciliği gibi teknolojilerin yükselişi, telif hukukuna zorluklarla birlikte fırsatlar da getirmiştir. Telif hukuku artık bu teknolojik gelişmelerin getirdiği karmaşıklıkları da ele almaktadır. Bu teknolojiler sadece yaratıcı eserleri kullanmanın yenilikçi yollarını sunmakla kalmayıp aynı zamanda telif hakkı olanaklarının kullanım kapsamını da genişletmektedir. Böylece telif hakkı, teknolojik gelişmelere uyum sağlamak için getirilen yeni kanuni düzenlemelerle fikri mülkiyet hukukunu sürekli olarak şekillendirmektedir. Dünyanın dört bir yanında metin ve veri madenciliği ekseninde yapılan, gerek telif haklarına istisna getirme şeklindeki düzenlemeler, gerekse bu uygulamaların adil kullanım çerçevesinde değerlendirilmesi şeklindeki düzenlemeler, hukuku bu gelişen teknolojilerin dinamik doğasıyla uyumlu tutmak için süregelen mücadeleyi temsil etmektedir. Bu çalışma, 2019 yılında Avrupa Birliği (AB) tarafından kabul edilen Dijital Tek Pazarda Telif Hakları Direktifinde (CDSM Direktifi) belirtilen metin ve veri madenciliği istisnalarına odaklanarak, bu düzenlemelerle teknolojik realitenin arasındaki uyumu incelemeyi amaçlamaktadır. Bu doğrultuda bu çalışma, yeni düzenlemelerin, küresel rekabet koşullarında Avrupa Birliği'ndeki inovasyon üzerindeki etkisini araştırmayı hedeflemektedir. Özetle, AB içindeki metin ve veri madenciliği istisnasının kapsamlı bir değerlendirmesi yapılmakta, Almanya ve Fransa'nın CDSM Direktifini iç hukuka aktarım süreçleri üzerinden metin ve veri madenciliği yaklaşımları incelenmektedir. Ardından daha esnek hukuki düzenlemelerin de değerlendirilebilmesi için Amerika Birleşik Devletleri ve Japonya'daki metin ve veri madenciliğine telif hakları

bakımından yaklaşımlar incelenmektedir. Son olarak konuya ilişkin düzenlemelerin henüz mevcut olmadığı Türkiye'nin hukuki ortamı incelenerek durumun kapsamlı bir şekilde anlaşılması amaçlanmaktadır.

Bu çalışma, yedi bölümden oluşmaktadır. **İlk bölüm**, metin ve veri madenciliği merkezinde yer alan temel kavramları tanıtmakta ve sonraki bölümlerde yapılacak tartışmalara zemin hazırlamaktadır. Bu incelemeler üç adım testi, metin ve veri madenciliği, yapay zekâ ve aralarındaki ilişki üzerinde yapılmıştır. Bu incelemelerde, kısaca büyük veri, veri koruma, veri kullanma ve dolayısıyla metin ve veri madenciliği dinamikleri açıklanmıştır. Büyük veriye bağlı bu dinamikler çalışma açısından önem arz etmektedir. Yapay zekâ kaynaklı yaratıcılıktaki artış, yapay zekâ modellerini eğitmek için bir ön koşul olan bu veri bolluğuna karmaşık bir şekilde bağlıdır. Bu çerçevede, metin ve veri madenciliği, yapay zekâ gelişimi için kritik olan yeni kalıpları ortaya çıkarmak için geniş dijital metin ve verilerin analiz edilmesini sağlayan vazgeçilmez bir araç olarak ortaya çıkmaktadır. Metin ve veri madenciliğinin en önemli özelliği, sıradan insan analizinin ötesine geçen veri kümelerindeki korelasyonları ortaya çıkarma yeteneğinde yatmaktadır. Teknik olarak sistemin işleyişini kısaca özetlemek gerekirse, otomatik araçlar kullanarak büyük miktarda dijital verinin tanımlanması, kopyalanması, çıkarılması ve yeniden birleştirilmesini içeren bir araştırma tekniğidir. Fen bilimlerinde kullanımı 1986 yılına kadar uzanmakta olup, Raynaud Hastalığı için balık yağının faydaları bu sayede keşfedilmiş ve 2005'te Huntington Hastalığı ile bağlantılı moleküller bu sayede tanımlanmıştır. Bu örneklerden de görülebileceği üzere bu sistematik süreç sayesinde büyük veriden insan eliyle yapmanın mümkün olmadığı keşifler yapılmakta, dolayısıyla yapay zekâ ve makine öğrenmesinde kritik ve yeri doldurulmaz bir rol oynamaktadır. Metin ve veri madenciliği gelişerek sektörleri değiştirmekte ve bilim ve teknolojiye yeni atılımlar için büyük bir potansiyel sunmaktadır.

İkinci bölüm metin ve veri madenciliğinin AB'de hangi koşullar altında düzenlendiğini incelemektedir. Bu bağlamda öncelikle uluslararası anlaşmalar ile kabul edilegelmiş telif hakkının koruma çerçevesinin metin ve

veri madenciliğini kapsayıp kapsamayacağı tartışılmıştır. Telif hakkı geleneksel olarak fikirler, metotlar, yalın gerçekler ve verilerden ziyade kavramların benzersiz ve yaratıcı biçimde ifade edilmesini koruduğu için metin ve veri madenciliği üzerindeki herhangi bir kısıtlama temel telif hakkı ilkelerini zayıflatıyor olarak algılanabilir. Bu sebeple, metin ve veri madenciliği faaliyetlerinin telif koruması kapsamına girmeyeceği yönünde görüşler bulunmaktadır. Keza bu görüşlere göre metin ve veri madenciliğine getirilecek sınırlamaların, Avrupa İnsan Hakları Sözleşmesi ve Avrupa Birliği Temel Haklar Bildirgesinde tanınan ifade ve bilgi özgürlüğü haklarını da sınırlandıracağı düşünülmektedir. Bu görüşlere karşı, çeşitli metin türlerinde veya yapılandırılmış/yapılandırılmamış veri kümelerinde sıklıkla bulunan korunmasız fikirlerin madencilik yapılması suretiyle yeni bir form almasının telif hakkı kapsamında değerlendirilmesi gerektiğini ileri süren görüşler de bulunmaktadır. Nitekim bu görüşler doğrultusunda getirilen telif hukukunda metin ve veri madenciliği istisnası, AB örneğinde olduğu gibi yalın gerçeklerin ve verilerin kullanımını yönetmeyi amaçlayarak telif hukukuna yeni bir boyut getirmektedir.

Her ne kadar metin ve veri madenciliği kendi başına telif hakkı kapsamında olmasa da metin ve veri madenciliği faaliyetleri sırasındaki belirli eylemler potansiyel olarak fikri mülkiyet haklarını ihlal edebilir. Bu sebeple AB, Dijital Tek Pazarı bütün üye ülkeler nezdinde uyumlu hale getirmek ve öncesinde parça parça düzenlenmiş olan AB seviyesindeki telif hukukunu düzenlemek amacıyla CDSM Direktifinde metin ve veri madenciliği istisnaları düzenlenmiştir. Metin ve veri madenciliği ile ilgili önceki yasal çerçeve, 2001/29/EC sayılı Direktif (InfoSoc Direktifi) ve 1996/9/EC sayılı Direktif (Veri tabanı Direktifi) tarafından verilen haklara dair istisnalardır. Metin ve veri madenciliği sürecinde koruma altındaki eserlerin kopyalarının oluşturulması gibi çoğaltmaları içeren faaliyetler, Infosoc Direktifinde açıklanan korumayı tetikleyebilir. Madencilik aşaması ise kullanılan yazılım ve çıkartmalara (extraction) bağlı olarak, Veri tabanı Direktifinde belirtildiği üzere çoğaltma hakkını da ihlal edebilir. Getirilen yeni istisnaların mahiyetini karşılaştırmak için bu yasal çerçeve de ikinci bölümde incelenmiştir.

Üçüncü bölümde düzenleyici çerçeve, yani CDSM Direktifinde metin ve veri madenciliği istisnalarının düzenlendiği 3'üncü ve 4'üncü maddeleri incelenmiştir. Madde 3'ün temel amacı, AB telif hakkı yasası çerçevesinde zorunlu bir istisna oluşturmaktır. Bu istisna, özellikle araştırma kuruluşları ve kültürel miras kurumları tarafından hali hazırda yasal erişimleri olan eserlerin veya diğer konuların metin ve veri madenciliğini gerçekleştirmek için yapılan çoğaltmalar ve çıkarmalara yöneliktir. Bu kapsamda, yapılmış metin ve veri madenciliğine konu eserlerin veya diğer konuların kopyaları uygun bir güvenlik seviyesinde saklanabilmesi ve araştırma sonuçlarının doğrulanması da dahil olmak üzere bilimsel araştırma amacıyla saklanabilmesi de öngörülmüştür. Bu istisnanın amacı, bilimsel araştırma amacıyla metin ve veri madenciliği faaliyetleri yürütmeyi kolaylaştırmaktır. Bu çerçevede bilimsel araştırmanın kapsamının hem fen hem de sosyal bilimlerinin içerdiği belirtilmiştir. Fikri mülkiyet haklarına saygı gösterirken bilimsel araştırmaları teşvik etmenin önemini kabul eden bu madde hem telif hakkı hem de sui generis veri tabanı hakkı için çok önemli bir istisna oluşturmaktadır.

Öte yandan, Madde 4, hem telif hakkı hem de sui generis veri tabanı hakkına ikinci bir istisna getirerek ticari anlamda da metin ve veri madenciliği faaliyetlerini teşvik etmektedir. Bu hüküm, metin ve veri madenciliği amaçları için yasal olarak erişilebilir içeriğin çoğaltılmasına ve çıkarılmasına izin vermektedir. Ancak Madde 3'ten farklı olarak, ticari araştırma, gazetecilik, veri analizi ve yapay zekâ dahil olmak üzere çeşitli amaçları kapsamaktadır. Bununla birlikte, Madde 4 önemli bir farkı, hak sahiplerine bu istisnayı açıkça saklı tutma seçeneğini vermekte ve devre dışı bırakma veya sözleşmeden çıkma gibi mekanizmalar yoluyla potansiyel olarak bu istisnanın geçersiz kılınmasına izin vermektedir. Başka bir deyişle, bu istisna yalnızca telif hakkı sahiplerinin bu kullanımı sözleşmeler, beyanlar, makine tarafından okunabilir araçlar veya hüküm ve koşullar yoluyla açıkça yasaklamamış olması durumunda mümkündür. Bu saklı tutma hakkı, metin ve veri madenciliği faaliyetlerinin telif hakkı sahiplerinin izin verip vermeme tercihine bağlı olduğunu ve aynı zamanda veri kullanımı için ücret talep edebilecekleri anlamına gelmektedir.

Dördüncü bölüm, metin ve veri madenciliği istisnaları ile ilgili eleştirileri ve zorlukları eleştirel bir şekilde analiz etmekte ve özellikle teknolojik gelişmelerin kısıtlanması potansiyeline odaklanmaktadır. Bu çerçevede Direktifin 3'üncü ve 4'üncü Maddeleri, metin ve veri madenciliğinin tanımı, yararlanıcılar, haklar, teknolojik adaptasyon kapasitesi ve birincil kaynaklara erişim gibi farklı açılardan değerlendirildiğinde sorunlar olduğu belirtilmektedir. Öncelikle metin ve veri madenciliği tanımı, Madde 2'de "kalıplar, eğilimler ve korelasyonlar gibi bilgiler üretmek için dijital formdaki metin ve verileri analiz etmeyi amaçlayan herhangi bir otomatik analitik teknik" olarak özetlenmiş ve Direktifin 8'inci dibacesinde "yeni teknolojiler tarafından etkinleştirilen metin, ses, görüntü veya veri gibi dijital formdaki bilgilerin otomatik hesaplamalı analizi" olarak daha da genişletilmiştir. Tanımla ilgili olarak dile getirilen endişeler, çok çeşitli otomatik analitik teknikleri kapsamasının çok geniş olabileceği yönündedir. Direktifte yapılan metin ve veri madenciliği tanımı, esasen yapay zekâ kaynaklı makine öğrenme sistemleri tarafından gerçekleştirilen ortak bir görevdir. Dolayısıyla, Direktifin lafzı dikkate alındığında, veri analizine dayandıkları sürece hem şu anda tanınan hem de gelecekte ortaya çıkacak yapay zekâ ve makine öğrenimi uygulamalarının faaliyetleri üzerinde bir engel olabileceği düşünülmektedir.

Ayrıca Direktifte tanınan metin ve veri madenciliği istisnasında bu istisnadan kimlerin yararlanabileceği açık bir şekilde sınırlandırılmış olup söz konusu kurumlarda çalışmayan bilim insanları ve bağımsız araştırmacılar dahil edilmemiştir. Bu sınırlama, metin ve veri madenciliği sayesinde erişilebilecek bilgiye eşit şartlarda erişimi engelleyebilir ve aynı şekilde yapay zekâ odaklı çalışan, üretim yapan veya hizmet veren yeni girişimler için bir engel niteliği taşıyabilir. Bu çerçevede 3'üncü maddenin yararlanıcılarının sınırlı tutulmuş olmasının yanı sıra 4'üncü maddede hak sahiplerine tanınan kapsam dışı bırakma mekanizması, AB'de yapay zekanın ilerlemesi üzerindeki potansiyel olumsuz etkisi nedeniyle güçlü eleştirilere maruz kalmıştır. Dördüncü maddede tanınan ticari istisnanın yapay zekâyı geliştirme ve araştırmalarını kolaylaştırmak için tasarlandığı göz önünde

bulundurulduğunda, hak sahiplerine belirli kullanımları kısıtlama yetkisi verilmesi, yasa koyucunun hak sahiplerinin menfaatleri ile ticari amaçla metin ve veri madenciliği faaliyeti yapmak isteyenler arasında kurmayı amaçladığı denge olarak yorumlanmaktadır. Bu hükümlerin fikri mülkiyet haklarını korurken veri blokajı riskini yönetmedeki etkinliği, yapay zekâ kaynaklı teknolojik gelişmelerle birlikte gözlemlenmesi gereken çok önemli bir husus olarak değerlendirilmiştir.

Söz konusu maddelere ilişkin eleştiriler incelendikten sonra teorik düzlemdeki bu eleştirilerin fiili uygulamadaki yansımalarını inceleyebilmek amacıyla çalışmanın **beşinci bölümünde**, AB’de öncü olarak kabul edilen iki önemli AB üye ülkesi olan Almanya ve Fransa’daki yasal mevzuata adaptasyon süreçlerini incelenmektedir. Bu çerçevede, Alman hukukunda metin ve veri madenciliğine getirilen yeni istisna ve sınırlama, özellikle Alman içtihadındaki “Freier Werkgenuss” doktrini, lafzi çeviriyle “çalışmalardan özgürce yararlanma” anlamına gelen, mahkeme içtihatlarıyla oluşmuş bu doktrin göz önünde bulundurulduğunda ilgi çekici bir örnek sunmaktadır. Japonya’nın metin ve veri madenciliğine yaklaşımına çok benzediği tespit edilen bu doktrin, telif haklarının var olmasının ana nedeni olan entelektüel veya sanatsal eserlerden aldığımız keyif ve tatmin etrafında dönmektedir. Metin ve veri madenciliğinin telif hakkı kurallarına sıkı sıkıya bağlı olmaması gerektiği, çünkü telif hakkının temel amacının entelektüel tatmini desteklemek olduğu savunulmaktadır. Buna göre eğer bir eylem bu tür bir haza gerçekten katkıda bulunmuyorsa, telif hakkı ile o kadar da ilgili olmayabilir. Bu sebeple, yeni yasal düzenlemelerle öngörülen metin ve veri madenciliği istisnalarının gelecekteki hukuki uyumsuzluklarda “Freier Werkgenuss” doktrini doğrultusunda nasıl yorumlanacağı merak konusudur.

Aynı bölümde incelenen Fransa’da da Direktifin ilgili maddeleri gereğince metin ve veri madenciliği istisnası iç hukukta düzenlenmiştir. Bilimsel araştırmalar için öngörülen istisna, ticari amaçlar için getirilen istisna ve bu istisna ile hak sahiplerine tanınmış olan herhangi bir gerekçe göstermeksizin bu istisnadan vazgeçmelerine olanak tanıyan kapsam dışı

tutma hakkı birlikte düzenlenmiştir. Kapsam dışı tutma hakkı, hak sahipleri tarafından yoğun ilgi görmüş ve çeşitli kuruluşlar bu maddenin uygulanmasını başlatmıştır. Örneğin 427.500 sanatçıyı temsil eden, 96 milyona yakın esere erişimi olan Müzik Yazarları, Bestecileri ve Yayıncıları Derneği (SACEM, Société des Auteurs, Compositeurs et Éditeurs de Musique), yeni mevzuat kapsamında hak sahiplerini temsilen kapsam dışı bırakma hakkını kullanmaya karar verdiğini duyurmuştur. Bu durum, hak sahiplerinin açık rızası olmadan yapay zekâ eğitimi için veri kullanımında daha sıkı kontrollere doğru bir anlayışın hâkim olmaya başladığını ispat eder niteliktedir. Tüm bu değişiklikler, dijital ortamda telifli içeriğin korunmasının önemine ilişkin artan bir farkındalığa işaret etmekte ve verilere erişimin yapay zekâ geliştiricileri açısından zorlaştırıldığı ve muhtemelen pahalılaştırıldığını da ortaya koymaktadır.

Metin ve veri madenciliği faaliyetlerinin herhangi bir izin gerektiremediği ve belli şartlar altında olduğunda telif hakkını ihlal etmediği yaklaşımı bulunan ülkelerin incelenmesi bu çalışmanın esas sorusunu cevaplayabilmek için önem taşımaktadır. Bu çerçevede **altıncı bölümde** incelenen Amerika Birleşik Devletleri (ABD)'nin ve Japonya'nın aralarında olduğu bu ülkelerin yaklaşımı, teknolojinin yasal yönlerini düzenlemenin zorluğuna dayanmaktadır. Bu zorluk, teknolojik ilerlemelerin hızı ile yasaları uygulamaya koyma süreçlerinin nispeten yavaş olması arasındaki doğal uyumsuzluktan kaynaklanmaktadır. Bu anlamda, dikkatli taslak hazırlama ve hararetli tartışmalara rağmen kanunların yürürlüğe girmesi ve uygulanması için geçen süre zarfında yasal hükümler hızla güncelliğini yitirebilmektedir. Bütün bu hususlar göz önünde bulundurulduğunda, ABD ve Japonya hukuk sistemi ayrıca önem taşımaktadır. Farklı telif hakkı geleneklerine ait olmalarına rağmen, her ikisi de yapay zekâ alanında güçlü yaratıcı endüstriler ve rekabetçi yüksek teknoloji sektörlerine engelsiz hukuki altyapı sağlamaya çalışmaktadır. Telif hakkı alanında, özellikle metin ve veri madenciliği ve makine öğrenmesi faaliyetleri ile ilgili olarak ABD, AB'ye kıyasla telif hakkıyla korunan materyallerle metin ve veri madenciliği faaliyetlerine izin verme konusunda "adil kullanım" doktrini ile daha esnek

olmasıyla bilinmektedir. Adil kullanım, telif hakkı sahibine izin veya tazminat ödenmesini gerektirmeden telif hakkıyla korunan materyallerin belirli kullanımlarına izin veren yasal bir ilkedir. İçtihatlar ve uygulama ile gelişen bu doktrin, telif hakkı sahibinin hakları ile kullanıcıların telif hakkıyla korunan materyalleri araştırmasına yönelik kamu yararı gibi belirli amaçlar için kullanma hakları arasında bir denge kurmaya çalışır. Sonuç olarak bu yaklaşım ABD’yi yapay zekâ araçları geliştirmek isteyen şirketler için cazip bir seçenek haline getirmektedir.

Aynı bölüm altında incelenen Japon Telif Kanunu, metin ve veri madenciliği faaliyetlerine “zevk almama” kistasını esas alan yeni bir bakış açısı getirmektedir. Geleneksel açıdan ele alındığında telif hukuku müzik veya edebiyat gibi entelektüel veya duygusal zevk veren eserleri korur, oysa bu görüşe göre metin ve veri madenciliği zevk alma amacı gütmeyen bir bilgisayar programıdır. Dolayısıyla doğal olarak telif yasasının korumayı amaçladığı menfaatlere zarar vermez. Japonya’nın metin ve veri madenciliğine eğlence dışı amaçlarla izin verme kararı, teknolojik ilerleme ile telif hakkı sahibinin haklarını dengelemeyi amaçlayan, telif hakkı ilkelerine ilişkin incelikli ve sıra dışı bir anlayışı yansıtmaktadır. Sonuç olarak yapay zekâ odaklı iş yürüten veya hizmet sunan şirketlere veya bireylere kolaylık sağladığı açıktır.

Çalışmanın **yedinci bölümünde**, karşılaştırmanın kapsamı metin ve veri madenciliği konusunda hukuki bir düzenleme barındırmayan Türkiye’nin yasal düzenlemeleri incelenerek genişletilmektedir. Bu bölümde yapılan inceleme ile görülmektedir ki, Türk hukukunda metin ve veri madenciliğine ilişkin istisnalara karşılık gelen bir hüküm bulunmamakta ve bu konuda görüşülmekte olan bir kanun tasarısı da bulunmamaktadır. 25 yıldır AB adayı olan Türkiye, AB müktesebatına uyum fasılları hakkında her yıl raporlar almaktadır. Bu bağlamda, AB Komisyonu’nun 2023 raporunda, Türkiye’deki mevcut telif yasasının teknolojik gelişmelere uygun olarak yeterli koruma sağlayacak şekilde tam bir modernizasyondan geçmediği ve Türkiye’nin CDSM Direktifi ile henüz uyumlu olmadığı belirtilmektedir. Bu

çerçeve, yürütülen metin ve veri madenciliği faaliyetlerinin mevcut mevzuat uyarınca hak sahiplerinin çeşitli haklarını ihlal edeceği kaçınılmazdır. Doktrin, bu faaliyetlere katılanların çıkarlarını, daha geniş kamu yararını ve eserleri bu süreçlerde kullanılan eser sahiplerinin haklarını dengelemek için yasal bir çerçeveye duyulan ihtiyacı kabul etmektedir.

Sonuç kısmında tez boyunca elde edilen bulguları ve eleştirileri sentezleyerek araştırma sorusunun yanıtlanması amacıyla konuya ilişkin nihai bir bakış açısı sunulması amaçlanmıştır. Bu çerçevede varılan sonuç, daha esnek hukuk sistemleri ile karşılaştırıldığında da açık olduğu gibi, metin ve veri madenciliği istisnalarının teknolojik gelişmeleri kısıtlayıcı nitelikte olduğu yönündedir. Direktifinin AB tarafından kabul edilmesinin hemen ardından Generative AI denilen üretken yapay zekâ sistemlerinde yaşanan gelişmeler yasal düzenlemelerin teknolojinin gerisinde kalmasına sebep olmuştur.

Anahtar Kelimeler: Metin ve veri madenciliği, Dijital Tek Pazarda Telif Hakları Direktifi, Adil kullanım doktrini, Telif Hakları istisnaları, Yapay zekâ hukuku

ABSTRACT

A Challenge to Copyright: Text and Data Mining

DOĞAN, Esma Muheyne

Master's Thesis, Department of Private Law

Supervisor: Assoc.Prof. Dr. Cahit SULUK

May 2024

The rise of technologies like Artificial Intelligence (AI), Machine Learning (ML) and subsequently Text and Data Mining (TDM), has brought both challenges and opportunities to the copyright law. Copyright law is now addresses the complexities introduced by these technological advancements. These technologies not only offer innovative ways to use creative works but also extend the scope of exercise of copyright possibilities. The complex interaction of copyright, technology, and societal norms, especially in the digital and intelligent systems context, is consistently molding the landscape of intellectual property. TDM regulations all around the world represents the ongoing struggle to keep regulations aligned with the dynamic nature of these evolving technologies. This thesis focuses on the current challenges presented by these advancements, with a specific emphasis on the TDM exceptions outlined in the Copyright in the Digital Single Market Directive (CDSM Directive) which adopted by European Union (EU) in 2019. This thesis aims to explore the impact of the TDM exception within the EU on innovation while investigating how the current state of the TDM exception may impose limitations on innovation in the EU. To address this, the study conducts a comprehensive evaluation of the TDM exception within the EU. In this perspective, TDM approaches of Germany and France through their transposition processes, TDM approaches in the United States and Japan for exploring the more flexible frameworks. And the regulatory landscape of Türkiye, where efforts on the subject are currently absent are examined to contribute a comprehensive understanding of the situation.

Keywords: Text and Data Mining, Copyright in the Digital Single Market Directive, Fair Use Doctrine, Non-Enjoyment Doctrine, Copyright Exception

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Cahit SULUK, for guiding me through this study. His thoughtful comments and recommendations have been invaluable to this dissertation.

I would also like to thank the jury members, Asst. Prof. Dr. Tuğba GÜLEŞ and Asst. Prof. Dr. Abdurrahman Hamza TÜZGEN, for their generous expertise and constructive feedback.

Lastly, I am deeply grateful to my parents for their unwavering support, and to my beloved siblings Ayşenur, Heysem, Sümel, Talha, and my dear friend Büşra Sultan, who stood by my side and encouraged me through the highs and lows of my academic journey.

Table of Contents

Abbreviations.....	19
Introduction	1
<i>Focus and Purpose</i>	2
<i>Methodology and Structure</i>	4
Chapter One: Fundamental Concepts	6
1.1- <i>The Three-step test</i>	8
1.2- <i>Text and Data Mining</i>	11
1.3- <i>Artificial Intelligence</i>	14
1.4- <i>The relationship between AI and TDM</i>	20
Chapter Two: Background of the situation of TDM	22
2.1- <i>Sui Generis Database Right</i>	25
2.2- <i>Information Society Directive</i>	32
2.3- <i>Infopaq Cases</i>	33
Chapter Three: The TDM exceptions	35
3.1- <i>Article 3 – Scientific Research Exception</i>	38
3.2- <i>Article 4 - General Exception</i>	39
Chapter Four: Assessment of the Articles	41
4.1- <i>Challenges with TDM Definitions and Beneficiaries</i>	42
4.2 - <i>TDM's Role in AI and Copyright Law</i>	43
4.3 - <i>Implications on Copyright Principles</i>	44
4.4 - <i>Shortcomings of the CDSM Directive</i>	45
4.5- <i>Assessment of Article 3</i>	45
4.6 - <i>Assessment of Article 4</i>	48
4.6.1- <i>The Technical Implementation of Article 4 Section 3</i>	51
Chapter Five: The Implementation of the Articles.....	55
5.1- <i>Germany</i>	55
5.2- <i>France</i>	56
Chapter Six: The regime of the TDM outside of the EU - United States and Japan	60
6.1- <i>Legal Framework in the United States</i>	60
6.2- <i>Legal Framework in Japan</i>	63

Chapter Seven: Legal Framework in Türkiye	65
<i>7.1- Current Situation</i>	65
<i>7.2- Legislative Background</i>	66
<i>7.3 - Challenges and Considerations in TDM and Legal Framework in Türkiye</i>	74
Conclusion.....	78
References	81

Abbreviations

AI	Artificial Intelligence
Art.	Article
CDSM Directive	Copyright in the Digital Single Market Directive
CJEU	Court of Justice of the European Union
EU	European Union
GenAI	Generative Artificial Intelligence
IoT	Internet of Things
IP	Intellectual Property
LIA	Law on Intellectual and Artistic Works
ML	Machine Learning
SMEs	Small and Medium sized Enterprises
TDM	Text and Data Mining
TRIPs Agreement	The Agreement on Trade-Related Aspects of Intellectual Property Rights
WIPO	World Intellectual Property Organization

Introduction

As technology has advanced, copyright law has undergone significant transformations. Initially designed for professional and commercial use, exclusive rights held a central position.¹ However, the copyright landscape has dynamically evolved, influenced by technology, changing social norms, and shifts in the market.² Innovations like TDM and ML systems have not only introduced novel ways of utilizing works but have also expanded the scope of copyright possibilities.³ The intricate interaction among copyright, technology, and societal norms in the context of digital progress contributes to the continuous development of copyright law.

Digital technologies are reshaping the creation, sharing, and use of creative content, presenting both opportunities and challenges. This affects various groups like authors, artists, educators, and online services. The global nature of digital networks raises questions about how well copyright rules are working together all around the world.⁴ Apart from digital aspect, the harmonization of copyrights' core economic rights at the international level remains limited, as seen in treaties such as the Berne Convention, TRIPs Agreement, and WIPO Copyright Treaty.⁵ These agreements grant contracting states substantial autonomy in implementing and interpreting economic rights, resulting in divergent definitions and interpretations, as can be seen different implementations on a minor issue like TDM all over the world. Thus, preceding the

¹ World Intellectual Property Organization, *Understanding Copyright and Related Rights*, second edition, 2016, 4, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.

² João Pedro de Miranda Branco Tomé Quintais, "Copyright in the Age of Online Access: Alternative Compensation Systems in EU Copyright Law" (phd Thesis, Universiteit van Amsterdam, 2017), 309, https://pure.uva.nl/ws/files/7607912/QUINTAIS_Doctoral_Dissertation_Final_Manuscript.pdf.

³ Péter Kollár, "Mind If I Mine? A Study on the Justification and Sufficiency of Text and Data Mining Exceptions in the European Union" (LLM Thesis, Maastricht, Maastricht university, 2021), 7, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3960570.

⁴ European Commission, "Impact Assessment on the Modernisation of EU Copyright Rules," Pub. L. No. SWD(2016) 301 (2016), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52016SC0301>.

⁵ P. Bernt Hugenholtz and Martin Kretschmer, *Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change Chapter 1: Reconstructing Rights: Project Synthesis and Recommendations*, Information Law Series, volume 41 (Alphen aan den Rijn, The Netherlands: Wolters Kluwer, 2018), 2.

enactment of the recent Copyright Directive⁶ within the EU, Hugenholtz and Kretschmer stated that “*reconstructing copyright is a task not meant for the impatient or the faint-hearted; a substantial amount of work lies ahead.*”⁷

The intersection of TDM with the existing intellectual property framework in the EU encountered potential challenges. TDM activities, depending on the chosen technique, may trigger concerns related to copyright infringement, substantial extraction from databases, or even reproduction and adaptation of computer programs.⁸ Against this backdrop, the CDSM Directive, published in 2019 and set to be adapted in 2021, remains highly relevant today. With a focus on creating a Digital Single Market, as outlined in its third recital, it aims to adjust exceptions for the digital environment, make content more accessible, and establish an effective marketplace for copyright.⁹ This directive seeks to organize a thriving European market while fostering fairness in society for the realms of creations, culture, and information.¹⁰ The ongoing significance and relevance of the Directive are evident in its navigation of the complex intersections of technology, copyright, and the digital environment.

Focus and Purpose

This thesis revolves around the examination of TDM exceptions outlined in the CDSM Directive. Despite the passage of five years since the EU implemented the Directive, discussions regarding these exceptions persist, remaining extensive and

⁶ “Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC,” OJ L130/92 Copyright and related rights in the Digital Single Market § (2019).

⁷ Hugenholtz and Kretschmer, *Copyright Reconstructed*, 10.

⁸ Rossana Ducato and Alain Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” *European Intellectual Property Review*, CRIDES Working Paper Series no. 1/2021, February 1, 2021, 2, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829858.

⁹ Recital 3 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 1.

¹⁰ Severine Dusollier, “The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition,” *Common Market Law Review* 57, no. Issue 4 (August 1, 2020): 979–1030, <https://doi.org/10.54648/COLA2020713>.

prominent. The primary focus of this thesis is to critically analyze the TDM exception within the EU and evaluate its potential impact on innovation.

The research question guiding this thesis is how the current state of the TDM exceptions may pose limitations on innovation within the EU. The TDM exceptions officially entered into force on 7 June 2019, with a stipulated deadline for transposition into Member States' national law by 7 June 2021.¹¹ However, the majority of Member States have not met this deadline, therefore the European Commission has been taking actions for member states to fully implement the CDSM Directive. On 26 July 2021, just before the deadline for implementation, Commission asked 23 member states for updates on their progress, starting the infringement process.¹² On 19 May 2022, warnings been sent to 13 countries to speed up their efforts.¹³ Finally, on 15 February 2023, Commission referred Bulgaria, Denmark, Finland, Latvia, Poland, and Portugal to the Court of Justice of the European Union for not completing the implementation.¹⁴ Overall, this situation raised questions about the effective integration and enforcement of the exception across the EU.¹⁵

The purpose of this study is to delve into the nuances of the criticisms surrounding the TDM exception. The implementation of this exception has been met

¹¹ "Directive - 2019/790 - EN - Dsm - EUR-Lex," accessed April 23, 2024, <https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX%3A32019L0790>.

¹² "Copyright: Commission Calls on Member States to Comply with EU Rules on Copyright in the Digital Single Market," Text, European Commission, accessed June 6, 2024, https://ec.europa.eu/commission/presscorner/detail/en/MEX_21_3902.

¹³ "Copyright: MSs Urged to Fully Transpose EU Copyright Rules," Text, European Commission - European Commission, accessed June 6, 2024, https://ec.europa.eu/commission/presscorner/detail/EN/IP_22_2692.

¹⁴ "11 Member States Referred to the Court of Justice of the EU," Text, European Commission - European Commission, accessed January 10, 2024, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_704; See also: Ana Lazarova, "The Last in Line: Bulgaria Implements the CDSM Directive," Kluwer Copyright Blog, December 27, 2023, <https://copyrightblog.kluweriplaw.com/2023/12/27/the-last-in-line-bulgaria-implements-the-cdsm-directive/>.

¹⁵ Eleonora Rosati, "The DSM Directive Two Years On: Do Things Ever Get Easier?," *Iic; International Review of Industrial Property and Copyright Law* 52, no. 9 (2021): 1139, <https://doi.org/10.1007/s40319-021-01082-6>; "The European Commission Referred 11 Member States to the Court of Justice of the European Union for Failing to Fully Transpose EU Copyright Rules into National Law | Shaping Europe's Digital Future," February 15, 2023, <https://digital-strategy.ec.europa.eu/en/news/european-commission-referred-11-member-states-court-justice-european-union-failing-fully-transpose>.

with intense criticism both preceding and following its enactment. Critics have raised concerns about its perceived effects on the EU's competitiveness in information technologies, particularly when compared to the more flexible TDM legal frameworks found in other regions, such as the United States and Japan.

In addition to evaluating the TDM exception within the EU, this study aims to examine three distinct jurisdictions to provide a comprehensive understanding of TDM practices. The first group involves scrutinizing the TDM approaches of member states Germany and France, as their transposition processes offer insights into the research question. The second group focuses on exploring more flexible TDM approaches globally, specifically in the United States and Japan. Finally, the third group considers Türkiye, a jurisdiction where regulatory efforts on the subject are not currently in place. By examining these diverse contexts, this research seeks to contribute a nuanced perspective on the implications of TDM exceptions on innovation within the European context.

Methodology and Structure

The methodology employed in this study involves a comprehensive review and analysis of legal documents, scholarly articles, and relevant literature related to TDM exceptions within the CDSM Directive. This will provide a thorough understanding of the legislative framework, historical context, and evolving discourse surrounding TDM within the EU.

To assess the impact of the TDM exception on innovation, a comparative analysis will be conducted, examining the experiences of the EU alongside regions with more flexible TDM legal frameworks, such as the United States and Japan. Additionally, this study will delve into the transposition processes of member states within the EU, specifically focusing on their approaches to TDM. Moreover, it will extend its examination to Türkiye, a state that currently lacks active or passive regulatory efforts in the realm of TDM, offering a comprehensive understanding of diverse regulatory landscapes and their potential consequences for innovation in information technologies. This comparative approach aims to highlight contrasts in the regulatory landscape and potential consequences for innovation in information technologies.

Based on this methodology, the thesis consists of seven comprehensive chapters dedicated to addressing the research question and fulfilling the study's overarching purpose of investigating the potential impact of the TDM exception on innovation within the EU. The first chapter introduces fundamental concepts central to TDM, laying the groundwork for the ensuing discussions in the subsequent chapters. Chapter two delves into the conditions under which TDM is regulated in the EU, while chapter three provides an in-depth examination of the regulatory framework itself.

Chapter four critically analyzes the critiques and challenges associated with TDM exceptions, specifically focusing on their potential to restrict innovation. Moving forward, chapter five explores the transposition processes in two pivotal EU member states, Germany and France, recognized as pioneers in the EU. Expanding the comparative scope, the legal frameworks of the United States, Japan, and Türkiye will be scrutinized in the following chapters.

In the end of the study, the conclusion synthesizes the findings and insights obtained throughout the thesis, offering a conclusive perspective on the elements shaping the process of answering the research question.

Chapter One: Fundamental Concepts

The internet, filled with information and opportunities, offers both positive and negative scenarios. This duality arises from the fundamental nature of the internet as a vast repository of data. This vast repository of data, often called “big data”, defined by its five characteristics: volume, veracity, variety, value, and velocity.¹⁶ Suggesting that the extensive volume of the data, its’ quality, how fast can this data set might be collected, the diversity in the sources and the data itself creates a great value.¹⁷

In this sense, Facebook-Cambridge Analytica scandal represents a good example to understand what ‘dystopian’ nature of the big data might be: Surveillance. It was exposed on March 17, 2018, revealing that Cambridge Analytica, a company specializing in electoral campaigns, accessed data from over fifty million Facebook users.¹⁸ This data played a pivotal role in developing analytical tools for major campaigns, including the Presidential election that took place in United States in 2016 and the Brexit campaigns in the United Kingdom.¹⁹ The incident drew significant attention, sparking inquiries into the dynamics of big data, data protection, and TDM.

Beyond sheer quantity, big data possesses characteristics such as massive volume, high velocity, diversity, scope, resolution, relational nature, and flexibility.²⁰ The concept of big data described as the “new oil”²¹, which highlights a shift where individuals, once producers of data, now act more like consumers, contributing

¹⁶ Jenn Cano, “The V’s of Big Data Velocity Volume Value Variety and Veracity,” accessed June 17, 2024, <https://www.coursehero.com/file/65841733/The-Vs-of-Big-Data-Velocity-Volume-Value-Variety-and-Veracitypdf/>.

¹⁷ Cano; Daniel J. Gervais, “Exploring the Interfaces Between Big Data and Intellectual Property Law,” *SSRN Electronic Journal*, 2019, 4, <https://doi.org/10.2139/ssrn.3360344>.

¹⁸ “Data for Sale | YaleGlobal Online,” accessed November 30, 2023, <https://archive-yaleglobal.yale.edu/content/data-sale>.

¹⁹ Carole Cadwalladr and Emma Graham-Harrison, “Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach,” *The Guardian*, March 17, 2018, sec. News, <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>; Mustafa Aksu et al., *Yapay Zeka ve Hukuk*, ed. Mustafa Aksu, 1st ed. (İstanbul: Onikilevha Yayıncılık, 2023), 257.

²⁰ OECD, “Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by ‘Big Data,’” OECD Digital Economy Papers, vol. 222, OECD Digital Economy Papers, June 18, 2013, 11, <https://doi.org/10.1787/5k47zw3fcp43-en>.

²¹ World Economic Forum January 2011 In Collaboration with Bain & Company, Inc., “Personal Data: The Emergence of a New Asset Class,” January 2011, 5.

personal data that generates profits for corporations and shapes the cyber landscape.²² Big data goes beyond just promising information on everything; it equips us with tools to comprehend this information without predefined categories. Consequently, the increase in AI-driven creativity is intricately tied to this abundance of data, a prerequisite for training AI models.²³ In this framework, TDM emerges as an indispensable tool, enabling the analysis of vast digital text and data to unveil novel patterns critical for AI development.²⁴ Its' crucial feature lies in its' ability to uncover correlations within data sets that go beyond ordinary human analysis. In this sense, Big Data, AI and TDM closely interconnected and interdependent concepts. However, it is essential to recognize that different legal norms come into play when analyzing non-personal data protected under copyright law, such as texts or structured databases. These distinctions carry significant, yet often misunderstood, consequences.²⁵ Elaborating on these consequences, while TDM can greatly benefit society, applying copyright law to it can influence how data can be used, shared, and protected.²⁶ This inevitably would have an influence on innovation, privacy, and legal frameworks on a broad scale. And regulating these subjects in a narrow way proves that this impact isn't seemed to be understood well by public or policy makers.

Intellectual property has a paradox; it aims to promote knowledge while simultaneously restricting it.²⁷ This creates a conflict between public and private

²² Ellen Emilie Henriksen, "Big Data, Microtargeting, and Governmentality in Cyber- Times. The Case of the Facebook-Cambridge Analytica Data Scandal." (Master thesis, Oslo, University of Oslo, 2019), 54, <http://www.duo.uio.no/>; Ş.Barış Özçelik, "Yapay Zekanın Veri Koruma, Sorumluluk ve Fikri Mülkiyet Açısından Ortaya Çıkardığı Hukuki Gereksinimler," *Adalet Dergisi*, no. 66 (May 2021): 93.

²³ European Commission, "White Paper on Artificial Intelligence – A European Approach to Excellence and Trust" (Brussels, 2020), 1, https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf; Kristina Christensen, "A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence" 4, no. 2 (2021): 18.

²⁴ Martin Senftleben et al., "Ensuring the Visibility and Accessibility of European Creative Content on the World Market," *Journal of Intellectual Property, Information Technology and E-Commerce Law* Vol. 13, no. No.1 (2022): 70–71, <http://dx.doi.org/10.2139/ssrn.3785272>.

²⁵ Martin Kretschmer and Thomas Margoni, "Data Mining: Why the EU's Proposed Copyright Measures Get It Wrong," *The Conversation*, May 24, 2018, <http://theconversation.com/data-mining-why-the-eus-proposed-copyright-measures-get-it-wrong-96743>.

²⁶ Gervais, "Exploring the Interfaces Between Big Data and Intellectual Property Law," 5.

²⁷ Giancarlo F. Frosio, "Communia Final Report on the Digital Public Domain," March 31, 2011, 18, https://www.communia-association.org/wp-content/uploads/COMMUNIA_Final_Report.pdf.

interests. In this sense, exceptions and limitations to copyright are crucial because they allow using copyrighted works without permission for important public reasons and to protect fundamental rights like freedom of expression.²⁸ So, the paradox is about finding a balance between promoting knowledge and restricting it through intellectual property rules. It can be said that the reflection of this paradox today is the exceptions concerning TDM, since its ability to make sense of potential of the big data and impact on the AI and ML systems cannot be overlooked. Within this background, before delving into the legal framework concerning TDM, it is essential to explore key concepts, including the three-step test, TDM, AI, and their interconnections.

1.1- The Three-step test

The three-step test in copyright law is a crucial mechanism for maintaining balance between authors' exclusive rights and imposed limitations.²⁹ The three-step test, originates from Article 9(2) of the Berne Convention³⁰, afterwards identified in several international agreements; the Articles 10(1), 10(2) of the WIPO Copyright Treaty³¹, the Article 16(2) of the WIPO Performances and Phonograms Treaty³², Articles 13,17,26 (2) and 30 of the TRIPs Agreement.³³

²⁸ Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," *SSRN Electronic Journal*, 2018, 8, <https://doi.org/10.2139/ssrn.3160586>.

²⁹ Martin Senftleben, "Copyright, Limitations and the Three-Step Test. An Analysis of the Three-Step Test in International and EC Copyright Law" (Phd Thesis, Universiteit van Amsterdam, 2004), 5, https://pure.uva.nl/ws/files/3645247/37378_Thesis.pdf; Andrew F. Christie and Robin Wright, "A Comparative Analysis of the Three-Step Tests in International Treaties," *IIC - International Review of Intellectual Property and Competition Law* 45, no. 4 (June 2014): 3, <https://doi.org/10.1007/s40319-014-0202-2>; See also: Nikos Koutras and Haydn Rigby, "A Scientific Analysis of the Three-Step Test: Through the Lenses of International and Australian Laws," *Publishing Research Quarterly* 38, no. 3 (September 2022): 503–18, <https://doi.org/10.1007/s12109-022-09898-x>; See also the reflections in Turkish Law: Serap Helvacı and Eray Aksın Atar, "Fikir ve Sanat Eserleri Hukukunda Üç Aşamalı Test," *Marmara Üniversitesi Hukuk Fakültesi Hukuk Araştırmaları Dergisi* 22, no. 3 (December 16, 2016): 1289–1302.

³⁰ "Berne Convention," World Intellectual Property Organization (WIPO), accessed June 7, 2024, <https://www.wipo.int/wipolex/en/text/283698>.

³¹ "WIPO Copyright Treaty," n.d.

³² "WIPO Performances and Phonograms Treaty," World Intellectual Property Organization (WIPO), accessed June 7, 2024, <https://www.wipo.int/wipolex/en/text/295578>.

³³ "WTO | Intellectual Property (TRIPS) - Agreement Text - Contents," World Trade Organization, accessed June 7, 2024, https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm.

This test provides a framework for allowing specific unauthorized reproductions of copyrighted works. Adherence to this test requires reproductions to satisfy three criteria: they must not interfere with the standard use of the work, refrain from causing undue harm to the author's legitimate interests and occur only under specific circumstances.³⁴ It aims to find a balance between protecting the creator's rights and allowing some exceptions for certain cases. Hence, the evolving interpretation of the three-step test emphasizes accommodating public interest within a balanced framework. In this new interpretation for a measure to pass the test, it must avoid being overly broad, protect the substantive income potential of rightsholders, and prevent disproportional harm.³⁵ Scholars advocate for a comprehensive approach, encouraging a comprehensive assessment of all three components of the test.³⁶ This approach recognizes potential threats from excessive copyright protection to public interests in scientific progress, cultural development, and economic advancement, while ensuring fair compensation for copyright holder.

In this regard, the three-step test acts as a safeguard, ensuring that intellectual works, though not entirely unrestricted, can be used for socially valuable purposes.³⁷ Aligning national copyright laws with the three-step test is crucial due to its foundational role in international copyright law.³⁸ Therefore national legislators are

³⁴ See: Article 9(2) "(2) It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author." World Intellectual Property Organization, ed., *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)*, WIPO Publication, no. 615 (Geneva: World Intellectual Property Organization, 1978), 55.

³⁵ P. Bernt Hugenholtz and Ruth Okediji, "Conceiving an International Instrument on Limitations and Exceptions to Copyright," *SSRN Electronic Journal*, 2012, 3, <https://doi.org/10.2139/ssrn.2017629>.

³⁶ Hugenholtz and Okediji, 3; Reto M Hilty, Christophe Geiger, and Jonathan Griffiths, "Towards a Balanced Interpretation of the 'Three-Step Test' in Copyright Law.," *European Intellectual Property Review*, no. Vol.4 (2016): 492; Geiger, Frosio, and Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," 8.

³⁷ Senfleben, "Copyright, Limitations and the Three-Step Test. An Analysis of the Three-Step Test in International and EC Copyright Law," 35.

³⁸ Senfleben, 118.

forced to comply to the three-step test as the primary limitation set by international copyright law when establishing limitations.³⁹

From the point of TDM, the central question revolves around whether TDM copying can be regarded as a significant act of reproduction, considering the broad and flexible right granted in Article 9(1) of the Berne Convention. This provision allows authors the exclusive right to authorize reproduction “*in any manner or form*”.⁴⁰ If TDM copying isn’t initially seen as a significant act of reproduction, there’s no need for limitations or exceptions, eliminating the defense against alleged non-compliance with the three-step test.⁴¹ However, recognizing the need for flexibility, the international copyright community has retained autonomy in deciding on novel copying categories.⁴² Contrary to automatically extending Article 9(1) Berne Convention’s comprehensive scope to cover new categories, negotiators have excluded certain types, such as temporary acts, from the international right of reproduction.⁴³ This means that Article 9(1) Berne Convention doesn’t inherently cover TDM until an international consensus is reached.

Given the lack of thorough international discussion on the right of reproduction’s scope concerning TDM, various regulations exist at different levels of restriction. There is currently no global agreement on how to define the intersection of copyright and research rights, particularly in the realms of TDM, ML, and AI.⁴⁴ However, there’s a shift in the fragmented landscape of exceptions as the WIPO Standing Committee on Copyright discusses harmonizing global TDM through a proposed treaty.⁴⁵ Nevertheless, current TDM regulations in the EU, represent

³⁹ Senfleben, 118; See also the compatibility of Turkish Law with the three step test Helvacı and Atar, “Fikir ve Sanat Eserleri Hukukunda Üç Aşamalı Test,” 1295.

⁴⁰ World Intellectual Property Organization, *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)*, 54.

⁴¹ Martin Senfleben, “Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?,” *IIC - International Review of Intellectual Property and Competition Law* 53, no. 10 (November 2022): 1483, <https://doi.org/10.1007/s40319-022-01266-8>.

⁴² Senfleben, 1484.

⁴³ Senfleben, 1484.

⁴⁴ Sean Flynn et al., “Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action,” *SSRN Electronic Journal*, 2020, 2, <https://doi.org/10.2139/ssrn.3578819>.

⁴⁵ Sean M. Fiil-Flynn et al., “Legal Reform to Enhance Global Text and Data Mining Research,” *Science* 378, no. 6623 (December 2, 2022): 952, <https://doi.org/10.1126/science.add6124>.

voluntary domestic regimes that surpass mandated copyright protection.⁴⁶ Whereas the opposite level of representation can be seen in the “non-enjoyment” clause of Japanese Copyright Act’s Article 30-4 which allows the use of a work in different ways, as long as it’s not for personal enjoyment or causing others to enjoy it.⁴⁷ Article 30-4 ensures that using a work for “non-enjoyment” purposes doesn’t disrupt normal exploitation or unfairly harm the rights of the copyright holder which in the end makes it in line with the three - step test.⁴⁸ The further details of this perspective explained under the Chapter six.

1.2- Text and Data Mining

TDM is a research technique that involves identifying, copying, extracting, and recombining large amounts of digital data using automated tools.⁴⁹ It holds a crucial role in AI, carrying significant importance, since this systematic process is vital for extracting valuable insights from extensive digital resources, playing a critical role in advancing research and facilitating knowledge discovery in the realm of AI.⁵⁰ As TDM advances, it not only changes industries but also offers the potential for new breakthroughs in science and technology.⁵¹

The technical process of TDM consists of three steps.⁵² In the initial step, content accessibility serves as a foundation, where raw, target, and preprocessed data

⁴⁶ Senfleben, “Compliance of National TDM Rules with International Copyright Law,” 1494.

⁴⁷ Tatsuhiro Ueno, “The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes – Recent Amendment in Japan and Its Implication,” *GRUR International* 70, no. 2 (March 1, 2021): 148, <https://doi.org/10.1093/grurint/ikaa184>.

⁴⁸ Ueno, 150.

⁴⁹ Geiger, Frosio, and Bulayenko, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects,” 5.

⁵⁰ Thomas Margoni, “Computational Legal Methods: Text and Data Mining in Intellectual Property Research,” in *Handbook of Intellectual Property Research*, ed. Irene Calboli and Maria Lilla Montagnani, 1st ed. (Oxford University PressOxford, 2021), 487,

<https://doi.org/10.1093/oso/9780198826743.003.0032>. Thomas Margoni, “Computational Legal Methods: Text and Data Mining in Intellectual Property Research,” in *Handbook of Intellectual Property Research*, ed. Irene Calboli and Maria Lilla Montagnani, 1st ed. (Oxford University PressOxford, 2021), 487, <https://doi.org/10.1093/oso/9780198826743.003.0032>.

⁵¹ Margoni, 487.

⁵² Artha Dermawan, “Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn from the Japanese ‘Nonenjoyment’ Purposes?,” *The Journal of World Intellectual Property*, May 27, 2023, 4, <https://doi.org/10.1111/jwip.12285>.

intertwine. In the second step, extraction and/or copying of content, is pivotal in transforming raw and processed data into recognizable patterns.⁵³ The third step is the essence of TDM, which is uncovering correlations, patterns, and relationships within vast datasets. Crucially, TDM outcomes usually present these insights without including substantial or direct content from the original texts, images, or other expressive elements in the analyzed dataset.⁵⁴ This distinctive feature highlights the transformative nature of TDM, emphasizing its focus on distilling valuable information from data while minimizing direct reliance on the specific content within the dataset.⁵⁵ Step three represents the final method in majority of Generative AI (GenAI) models. Here, TDM's role extends into data cleaning, preprocessing, transformation, and pattern evaluation. This intricate relationship between TDM and AI illustrates a seamless flow, emphasizing the interdependent connection between data mining activities and the AI.⁵⁶

TDM have proven to be a versatile tool with applications spanning across diverse domains.⁵⁷ From linguistics and astronomy to the arts, TDM has showcased its potential to analyze extensive datasets, uncovering patterns and generating valuable insights.⁵⁸ In life sciences, TDM, dating back to 1986, uncovers hidden correlations.⁵⁹ For instance, researchers discovered fish oil benefits for Raynaud's Disease in 1986,

⁵³ Flynn et al., "Implementing User Rights for Research in the Field of Artificial Intelligence," 4.

⁵⁴ Michael W Carroll, "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful," *UC Davis Law Review- American University*, WCL Research Paper No. 2020-15, 53 (December 1, 2019): 895.

⁵⁵ Mark A. Lemley and Bryan Casey, "Fair Learning," *SSRN Electronic Journal*, 2020, 111, <https://doi.org/10.2139/ssrn.3528447>.

⁵⁶ Dermawan, "Text and Data Mining Exceptions in the Development of Generative AI Models," 4–6.

⁵⁷ Dusollier, "The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition," 984.

⁵⁸ For further research: Mehrdad Cheshmeh Sohrabi and Amir Mashhadi, "Using Data Mining, Text Mining, and Bibliometric Techniques to the Research Trends and Gaps in the Field of Language and Linguistics," *Journal of Psycholinguistic Research* 52, no. 2 (April 1, 2023): 607–30, <https://doi.org/10.1007/s10936-022-09911-6>; Haojing Chen, YP Tsang, and CH Wu, "When Text Mining Meets Science Mapping in the Bibliometric Analysis: A Review and Future Opportunities," *International Journal of Engineering Business Management* Volume 15 (December 2023), <https://doi.org/doi.org/10.1177/18479790231222349>.

⁵⁹ Matthew Sag, "The New Legal Landscape for Text Mining and Machine Learning," *Journal of the Copyright Society of the USA* Vol. 66 (February 27, 2020): 4, <http://dx.doi.org/10.2139/ssrn.3331606>.

and in 2005, molecules linked to Huntington's Disease were identified using TDM.⁶⁰ The Next Rembrandt project stands as a testament to TDM's impact on artistic endeavors, producing a remarkable portrait in the style of Rembrandt through meticulous analysis.⁶¹

The potential efficiency of TDM might be examined through the knowledge repetition in scholarly articles. According to this idea, scholarly article that publishes each year reaches to remarkable numbers that finding related one in this information overload is a serious problem for researchers.⁶² Statistics shows that half of these articles only read by people who obliged to read them, writers, editors and referees, and only %10 of them is cited.⁶³ Which means that scientific community may constantly reproduce the knowledge that already exists whereas certain discoveries doesn't come out in daylight. Subsequently, resources, research funds and time spent inefficiently in this process. In this sense from a technical perspective, TDM has the potential to address this issue by comprehensively scanning, processing, analyzing, and categorizing this vast knowledge in novel and unexplored ways. Hence, the importance of TDM shines through as it empowers scholars to navigate this vast sea of information effectively.⁶⁴ It serves as a guiding light, helping researchers pinpoint the articles most pertinent to their inquiries.

Moreover, TDM's influence extends beyond pure science, making significant strides in industrial and commercial sectors.⁶⁵ Its growing significance is evident in its widespread use across diverse sectors such as journalism, information retrieval, search engines, consumer information and pharmaceutical research.⁶⁶ For instance,

⁶⁰ Sag, 4.

⁶¹ "The Next Rembrandt: Recreating the Work of a Master with AI," accessed December 5, 2023, <https://news.microsoft.com/europe/features/next-rembrandt/>.

⁶² Elisabeth Pain, "How to Keep up with the Scientific Literature | Science | AAAS," accessed December 16, 2023, <https://www.science.org/content/article/how-keep-scientific-literature>.

⁶³ Thomas Margoni and Martin Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," *GRUR International* 71, no. 8 (2022): 689, <https://doi.org/10.1093/grurint/ikac054>.

⁶⁴ Carroll, "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful," 901.

⁶⁵ Otmame Azeroual, "Text and Data Quality Mining in CRIS," *Information* 10, no. 12 (December 2019): 374, <https://doi.org/10.3390/info10120374>.

⁶⁶ For further research: Shintaro Okazaki et al., "Using Twitter to Engage with Customers: A Data Mining Approach," *Internet Research* 25, no. 3 (June 1, 2015): 416–34,

integrating TDM into Smart Disclosure Systems addresses the common problem of “signing without reading” in online transactions.⁶⁷ Smart disclosure systems defined as sharing complex information in easy-to-read formats, helping consumers make informed choices.⁶⁸ They operate by employing TDM techniques like natural language processing and machine learning to automate the analysis of digital content, such as contracts and privacy policies, extracting information and identifying patterns.⁶⁹ In the end they make it simple for users to access important pre-contractual details and get personalized advice based on their preferences. TDM’s pivotal role in developing intelligent applications, particularly in tasks like translation and speech recognition, underlines its contribution to the evolving landscape of AI.⁷⁰ As we navigate the future, it is evident that TDM will continue to play a crucial role in driving advancements across various domains.

1.3- Artificial Intelligence

AI is a technology that combines data, algorithms, and computer power.⁷¹ It helps machines imitate human thinking for tasks such as recognizing faces, making decisions, predicting outcomes, and automating processes.⁷² With further developments, it became an overarching system that entails Machine Learning, Neural

<https://doi.org/10.1108/IntR-11-2013-0249>; Elham Kariri et al., “Exploring the Advancements and Future Research Directions of Artificial Neural Networks: A Text Mining Approach,” *Applied Sciences* 13, no. 5 (January 2023): 3186, <https://doi.org/10.3390/app13053186>.

⁶⁷ Rossana Ducato and Alain Strowel, “Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to ‘Machine Legibility,’” *IIC - International Review of Intellectual Property and Competition Law* 50, no. 6 (July 2019): 650, <https://doi.org/10.1007/s40319-019-00833-w>.

⁶⁸ “Informing Consumers through Smart Disclosure,” [whitehouse.gov](https://obamawhitehouse.archives.gov/blog/2012/03/30/informing-consumers-through-smart-disclosure), March 30, 2012, <https://obamawhitehouse.archives.gov/blog/2012/03/30/informing-consumers-through-smart-disclosure>.

⁶⁹ Ducato and Strowel, “Limitations to Text and Data Mining and Consumer Empowerment,” 658.

⁷⁰ P. Bernt Hugenholtz, “The New Copyright Directive: Text and Data Mining (Articles 3 and 4),” Kluwer Copyright Blog, July 24, 2019, <https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>.

⁷¹ European Commission, “White Paper on Artificial Intelligence – A European Approach to Excellence and Trust,” 2.

⁷² European Commission, 2.

Networks, Deep Learning, Generative AI and Large Language Models in given order.⁷³

ML, a pivotal component of AI, empowers machines to learn from numerous examples, constructing “mental” models for autonomous decision-making.⁷⁴ Using training algorithms, machines analyze organized training data sets, improving their performance over time without requiring specific programming.⁷⁵ It serves as a powerful tool for optimization, since its’ capacity to make predictions minimizes errors that would otherwise result from mere guesswork.⁷⁶ As a technique of automated data analysis, ML involves the deployment of TDM methods.⁷⁷ In this sense, the current rise of AI is attributed to the abundance of training data, advanced computational power, and ongoing innovations in ML algorithms in the big data-driven era.⁷⁸

Neural networks, a subset of ML and the basis of deep learning, named after mimicking brain neurons in signaling.⁷⁹ Google’s search algorithm is one of its most known examples.⁸⁰ Depending on neural networks, deep learning requires more data for accuracy, while ML functions effectively with less data.⁸¹

GenAI, rooting from deep learning, uses ML to craft new content mirroring training data. Unlike traditional ML, it prioritizes creating data resembling its training set, producing entirely new assets. The distinction is reflected in their evaluation metrics-machine learning relies on quantitative measures, while GenAI is qualitatively

⁷³ IBM Data and AI Team, “AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?,” *IBM Blog* (blog), July 6, 2023, <https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/>.

⁷⁴ Theodoros Chiou, “Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?,” 2019, 399, https://www.jipitec.eu/issues/jipitec-10-3-2019/5025/chiou_pdf.pdf.

⁷⁵ Harry Surden, “Artificial Intelligence and Law: An Overview,” *Georgia State University Law Review* 35, no. 4 (2019): 1311, <https://scholar.law.colorado.edu/cgi/viewcontent.cgi?article=2340&context=faculty-articles>.

⁷⁶ Team, “AI vs. Machine Learning vs. Deep Learning vs. Neural Networks.”

⁷⁷ Chiou, “Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?,” 399.

⁷⁸ Surden, “Artificial Intelligence and Law: An Overview,” 1316.

⁷⁹ “Explained: Neural Networks,” MIT News | Massachusetts Institute of Technology, April 14, 2017, <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.

⁸⁰ “How Does Google Search Work? The Google Algorithm Explained,” Intuji, February 3, 2023, <https://intuji.com/how-does-google-search-work/>.

⁸¹ Team, “AI vs. Machine Learning vs. Deep Learning vs. Neural Networks.”

assessed for realism, coherence, and diversity.⁸² Large language models, also originating from deep learning systems, are advanced computational models designed for understanding and generating human language.⁸³ With extensive parameter sizes and remarkable learning capacities, they excel in predicting word sequences and generating new text based on input.⁸⁴ One of the most prominent example of them would be ChatGPT.⁸⁵

GenAI systems, by design, mimic human creativity through the use of existing works of literature and art as training material.⁸⁶ The ML algorithms employed in this process analyze human works to recognize patterns and similarities, allowing the AI system to learn and imitate the style of these human creations. Essentially, the GenAI system becomes capable of producing novel works of literature and art on its own, drawing from a computational analysis of the diverse human works used in its training.⁸⁷

In this evolving digital landscape, the rapid growth of global data, projected to reach “175 zettabytes by 2025 from 33 zettabytes in 2018”, signals transformative opportunities.⁸⁸ Recognizing the fundamental role of data, enhancing access and management becomes imperative for the development of AI and other digital applications.⁸⁹ The close relationship between data and technology indicates how they work together to drive progress. Notably, the year 2022 has witnessed the evolution of

⁸² Dianne Castillo, “Generative AI vs. Machine Learning,” Seldon, November 10, 2023, <https://www.seldon.io/generative-ai-vs-machine-learning>.

⁸³ Yupeng Chang et al., “A Survey on Evaluation of Large Language Models” (arXiv, October 17, 2023), 4, <http://arxiv.org/abs/2307.03109>.

⁸⁴ Chang et al., 4.

⁸⁵ Mark Riedl, “A Very Gentle Introduction to Large Language Models without the Hype,” *Medium* (blog), May 25, 2023, <https://mark-riedl.medium.com/a-very-gentle-introduction-to-large-language-models-without-the-hype-5f67941fa59e>.

⁸⁶ “What Is ChatGPT, DALL-E, and Generative AI? | McKinsey,” accessed December 13, 2023, <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>.

⁸⁷ Martin Senftleben, “Generative AI and Author Remuneration,” *IIC - International Review of Intellectual Property and Competition Law* 54, no. 10 (November 1, 2023): 1538, <https://doi.org/10.1007/s40319-023-01399-4>.

⁸⁸ European Commission, “White Paper on Artificial Intelligence – A European Approach to Excellence and Trust,” 4.

⁸⁹ European Commission, 8.

GenAI models, marking a significant milestone in digital transformation.⁹⁰ However, the extensive adoption of GenAI tools has led to crucial conversations about their impact on intellectual property, specifically within the global copyright and patent framework.⁹¹ For instance, training AI with copyrighted works without requesting permission would result that a work is really similar to original work. Today this possibility is overly heard in the media, the examples of AI generating a really similar work can be seen in the cases of AI creating a new work after feeding and rendering a specific artist works.⁹² Again, AI used in restoring a Rembrandt's painting as well which proved remarkable outcome, given that it was cut off from all the edges to fit into a wall.⁹³ This result presumably might be considered as a copyright infringement.⁹⁴ Regardless of the legal actions might be taken by the owner concerning compensation, in the long run, this may have more drastic consequences, such as AI prevailing the art industry or even more consequences that can't be foreseen now.⁹⁵

⁹⁰ Maanak Gupta et al., "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy," *IEEE Access* 11 (2023): 1, <https://doi.org/10.1109/ACCESS.2023.3300381>.

⁹¹ Renana Peres et al., "On ChatGPT and beyond: How Generative Artificial Intelligence May Affect Research, Teaching, and Practice," *International Journal of Research in Marketing* 40, no. 2 (June 2023): 274, <https://doi.org/10.1016/j.ijresmar.2023.03.001>.

⁹² "The Next Rembrandt: Recreating the Work of a Master with AI"; Andy Baio, "Invasive Diffusion: How One Unwilling Illustrator Found Herself Turned into an AI Model," *Waxy.Org* (blog), November 1, 2022, <https://waxy.org/2022/11/invasive-diffusion-how-one-unwilling-illustrator-found-herself-turned-into-an-ai-model/>.

⁹³ "How AI Learned to Paint like Rembrandt," *Big Think* (blog), July 26, 2021, <https://bigthink.com/high-culture/ai-learned-paint-like-rembrandt/>.

⁹⁴ "Can an Artist Sue an AI over Copyright Infringement?," *CBC Radio*, October 12, 2018, <https://www.cbc.ca/radio/spark/409-1.4860495/can-an-artist-sue-an-ai-over-copyright-infringement-1.4860762>; Özçelik, "Yapay Zekanın Veri Koruma, Sorumluluk ve Fikri Mülkiyet Açısından Ortaya Çıkardığı Hukuki Gereksinimler," 109.

⁹⁵ Dani Di Placido, "The Problem With AI-Generated Art, Explained," *Forbes*, accessed June 12, 2024, <https://www.forbes.com/sites/danidiplacido/2023/12/30/ai-generated-art-was-a-mistake-and-heres-why/>; Jayeesha Deb, "AI Art Generates Controversy Among Artists," *The Arcadia Quill* (blog), accessed June 12, 2024, <https://arcadiaquill.com/25242/showcase/ai-art-generates-controversy-among-artists/>; "The First Piece of AI-Generated Art to Come to Auction," accessed June 12, 2024, <https://www.christies.com/en/stories/a-collaboration-between-two-artists-one-human-one-a-machine-0cd01f4e232f4279a525a446d60d4cd1>; "The Art World's AI Dilemma: Informed Insight from Industry Experts," *The Art Newspaper* - International art news and events, May 30, 2024, <https://www.theartnewspaper.com/2024/05/30/the-art-worlds-ai-dilemma-informed-insight-from-industry-experts>; "Sony World Photography Award 2023: Winner Refuses Award after Revealing AI Creation," April 17, 2023, <https://www.bbc.com/news/entertainment-arts->

Legal challenges, such as lawsuits from entities like Getty Images, highlight concerns about copyright infringements by AI companies, raising complex questions at the intersection of AI technology and intellectual property rights.⁹⁶ In this perspective, far-reaching consequences of AI technology for both science and society led to regulation attempts all over the world.⁹⁷ While AI developers caution that expansive and intrusive regulations may impede innovation,⁹⁸ the period from 2016 to 2022 has witnessed a significant surge in the total number of AI-related regulations enacted into law, rising from just 1 in 2016 to 37 in 2022.⁹⁹ Even though the regulations concerning AI is beyond the scope of this thesis, it should be stated that training data element of AI systems, which is equals to TDM, is the often-neglected

65296763; James Vincent, “The Scary Truth about AI Copyright Is Nobody Knows What Will Happen Next,” *The Verge*, November 15, 2022, <https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data>; See more: Andy Baio, “AI Data Laundering: How Academic and Nonprofit Researchers Shield Tech Companies from Accountability,” *Waxy.Org* (blog), September 30, 2022, <https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/>.

⁹⁶ Peres et al., “On ChatGPT and Beyond,” 274., See: “Getty Images sues AI Art Generator Stable Diffusion in the US for Copyright Infringement: <https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion>, See: Kyle Wiggers, “The Current Legal Cases against Generative AI Are Just the Beginning,” *TechCrunch* (blog), January 27, 2023, <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/>.

⁹⁷ “The World Wants to Regulate AI, but Does Not Quite Know How,” *The Economist*, accessed December 23, 2023, <https://www.economist.com/business/2023/10/24/the-world-wants-to-regulate-ai-but-does-not-quite-know-how>; “China and the Great Global AI Governance Divide | Lowy Institute,” accessed June 12, 2024, <https://www.loyyinstitute.org/the-interpreter/china-great-global-ai-governance-divide>; “AI Act | Shaping Europe’s Digital Future,” May 29, 2024, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>; The White House, “FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,” The White House, October 30, 2023, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>; “Hiroshima Process International Guiding Principles for Advanced AI System | Shaping Europe’s Digital Future,” October 30, 2023, <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system>.

⁹⁸ “The World Wants to Regulate AI, but Does Not Quite Know How.”

⁹⁹ Nestor Maslej et al., “The AI Index 2023 Annual Report” (Stanford, CA: Institute for Human-Centered AI, Stanford University, April 2023), 267, https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf.

aspect of AI discussions.¹⁰⁰ Despite the abundance of legal studies focusing on AI applications and outcomes, training data has received comparatively less attention.¹⁰¹ Acknowledging the crucial role of training data in ML techniques, it becomes evident that addressing regulatory issues related to training data is as, if not more, essential than regulating the algorithms themselves.¹⁰² In this context, the AI Act in the EU, adopted in 13 March 2024, explicitly associates the utilization of copyrighted works for training AI models with the TDM exception outlined in Article 4 of the CDSM Directive in its Recital 105.¹⁰³ This direct reference reinforces the applicability of the TDM exception to AI model training.¹⁰⁴ In conclusion, it can be said that CDSM Directive continues to shape the harmonized TDM exception, unaffected by the AI Act.¹⁰⁵ Yet concerns about hindering innovation may persist, as indicated by over 150 executives of major European companies letter to the EU, expressing worry about the proposed EU AI regulation, fearing it will adversely affect innovation and competitiveness.¹⁰⁶ They argue that stringent regulation, especially on GenAI, may

¹⁰⁰ Philipp Hacker, “A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act,” *Law, Innovation and Technology* 13, no. 2 (July 3, 2021): 259, <https://doi.org/10.1080/17579961.2021.1977219>.

¹⁰¹ Hacker, 258.

¹⁰² Hacker, 260.

¹⁰³ Recital 105: “*General-purpose AI models, in particular large generative AI models, capable of generating text, images, and other content, present unique innovation opportunities but also challenges to artists, authors, and other creators and the way their creative content is created, distributed, used and consumed. The development and training of such models require access to vast amounts of text, images, videos, and other data. Text and data mining techniques may be used extensively in this context for the retrieval and analysis of such content, which may be protected by copyright and related rights. Any use of copyright protected content requires the authorisation of the rightsholder concerned unless relevant copyright exceptions and limitations apply. Directive (EU) 2019/790 introduced exceptions and limitations allowing reproductions and extractions of works or other subject matter, for the purpose of text and data mining, under certain conditions. Under these rules, rightsholders may choose to reserve their rights over their works or other subject matter to prevent text and data mining, unless this is done for the purposes of scientific research. Where the rights to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightsholders if they want to carry out text and data mining over such works.*” “Recital 105 | EU Artificial Intelligence Act,” accessed June 12, 2024, <https://artificialintelligenceact.eu/recital/105/>.

¹⁰⁴ “A First Look at the Copyright Relevant Parts in the Final AI Act Compromise,” Open Future, accessed December 24, 2023, <https://openfuture.eu/blog/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise>.

¹⁰⁵ Matt Hervey, “The EU AI Act and IP,” LoupedIn, June 26, 2023, <https://loupedin.blog/2023/06/the-eu-ai-act-and-ip/>.

¹⁰⁶ Open Letter, “Open Letter to the Representatives of the European Commission, the European Council and the European Parliament,” accessed December 25, 2023, <https://www.igizmo.it/wp-content/uploads/2023/06/Open-Letter-EU-AI-Act-and->

lead to disproportionate compliance costs, increased liability risks, and companies departing from the EU.¹⁰⁷ This concern is shared among member states, with both the French and German governments actively investing in their domestic AI industries by allocating significant funds to support development.¹⁰⁸ While these governments are dedicated to promoting innovation, they have raised concerns that excessive regulation might impede their AI industries.¹⁰⁹ In this sense, French Finance Minister Bruno Le Maire said: “*Before regulating we must innovate.*”¹¹⁰ In light of these acknowledgments emphasizing the significance of prioritizing innovation in AI regulation, it can be argued that the TDM exceptions, in their regulation of AI training data, pose a potential contradiction by potentially impeding the very innovation they aim to foster. As further explained in Chapter 4, the exceptions are suitable for only limited beneficiaries for limited purposes which inevitably effects freely use of data by new AI technologies.

1.4- The relationship between AI and TDM

AI systems thrive on new data-driven technologies.¹¹¹ They rely on various types of data, like literary and artistic works (text, music, images) for their development.¹¹² The value of this data is vital for shaping the current and future capabilities of GenAI.¹¹³ The importance of data lies not just in its content but in the value extracted from it.¹¹⁴ To discover new patterns in creative outputs, GenAI needs

Signatories.pdf; Javier Espinoza, “European Companies Sound Alarm over Draft AI Law,” *Financial Times*, June 30, 2023, sec. EU tech regulation, <https://www.ft.com/content/9b72a5f4-a6d8-41aa-95b8-c75f0bc92465>.

¹⁰⁷ Open Letter, “Open Letter to the Representatives of the European Commission, the European Council and the European Parliament.”

¹⁰⁸ Will Henshall, “E.U.’s AI Regulation Could Be Softened After Pushback,” *TIME*, November 22, 2023, <https://time.com/6338602/eu-ai-regulation-foundation-models/>.

¹⁰⁹ Henshall.

¹¹⁰ Henshall.

¹¹¹ Andrej Karpathy et al., “Generative Models,” OpenAI.com, June 16, 2016, <https://openai.com/research/generative-models>.

¹¹² Senftleben et al., “Ensuring the Visibility and Accessibility of European Creative Content on the World Market,” 70; Aksu et al., *Yapay Zeka ve Hukuk*, 138–39.

¹¹³ Andrei-Dragoş Popescu, “The Value of Data From an Artificial Intelligence Perspective,” *Annals of the University of Craiova for Journalism, Communication and Management* Volume 5 (November 2019): 176.

¹¹⁴ Eleonora Rosati, “An EU Text and Data Mining Exception for the Few: Would It Make Sense?,” *Journal of Intellectual Property Law & Practice* 13, no. 6 (June 1, 2018): 429, <https://doi.org/10.1093/jiplp/jpy063>.

to analyze vast amounts of data. This analysis, too complex for manual methods, is efficiently done through an automated process known as TDM.¹¹⁵ Like any other type of ML, TDM technologies evolve by conducting numerous experiments using vast datasets.¹¹⁶ In essence, the close ties between GenAI, data, and TDM are crucial for expanding the creative possibilities of AI.¹¹⁷

The CDSM Directive outlines TDM as encompassing “*any automated analytical technique aimed at analyzing text and data in digital form to generate information, including patterns, trends, and correlations.*” This expansive definition inherently extends its reach to numerous training activities essential for the development of various AI systems, particularly those falling under the category of ML, such as GenAI systems.¹¹⁸ The broad scope outlined in the directive implies a comprehensive inclusion of activities crucial for advancing AI technologies.¹¹⁹

In this sense, copyright issue facing with GenAI involves tackling legal questions from two perspectives: the input side, dealing with the training activities crucial for building AI systems, and the output side, which raises concerns about the protection and potential infringement of AI-generated content.¹²⁰ This thesis focuses on the input side.¹²¹ In the EU, CDSM Directive addresses input-related issues through TDM exceptions in Articles 3 and 4, covering scientific and commercial TDM, respectively. Whereas in the United States, the absence of specific TDM exceptions prompts a fair use analysis for these activities which will be examined in the chapter six.

¹¹⁵ Eleonora Rosati, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects,” *Briefing Requested by the JURI Committee, Policy Department for Citizens’ Rights and Constitutional Affairs, European Parliament*, n.d., 2.

¹¹⁶ Carroll, “Copyright and the Progress of Science: Why Text and Data Mining Is Lawful,” 905.

¹¹⁷ Dermawan, “Text and Data Mining Exceptions in the Development of Generative AI Models,” 2.

¹¹⁸ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 686.

¹¹⁹ João Pedro Quintais, “Generative AI, Copyright and the AI Act,” Kluwer Copyright Blog, May 9, 2023, <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/>.

¹²⁰ Senfleben, “Generative AI and Author Remuneration,” 1542–49.

¹²¹ For a detailed research in Turkish on the output side of the AI: Mustafa Aksu, *Yapay Zeka ve Hukuk*, 1st ed. (İstanbul: Onikilevha Yayıncılık, 2023); Hasan Kadir Yılmaztekin, *Yapay Zekanın Eser Sahipliği* (Ankara: Adalet Yayınları, 2021).

Chapter Two: Background of the situation of TDM

While copyright traditionally protects the creative form rather than the underlying data, any restrictions on TDM could be perceived as undermining fundamental copyright principles.¹²² Regarding fundamental copyright principles, the legal framework safeguards the unique expression of concepts but does not extend protection to the concepts themselves, nor to plain facts or data.¹²³ Hence, TDM should be viewed as falling outside the boundaries of copyright infringement, as it involves matters beyond the scope of copyright.¹²⁴ Such limitations may also effect on the protected rights of freedom of expression and information, as recognized by entities like the European Court of Human Rights and the Charter of Fundamental Rights of the EU.¹²⁵ Consequently, introducing a copyright exception to govern the utilization of unprotected ideas, principles, facts, and data - frequently found in literary works or various text types (text mining) or structured/unstructured datasets (data mining) - poses challenges.¹²⁶

Widely used in scientific research, medicine, journalism, and AI as explained in previous chapters, TDM relies on making transient copies of data for analysis.¹²⁷ From a copyright perspective, this copying is considered reproduction, leading to debates about control and prior authorization by copyright holders.¹²⁸ Despite arguments that this copying is technical and not an act of exploitation, the EU opted

¹²² Geiger, Frosio, and Bulayenko, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects,” 4. , See: Fiil-Flynn et al., “Legal Reform to Enhance Global Text and Data Mining Research.”

¹²³ Eleonora Rosati, *Handbook on the Economics of Copyright: A Guide for Students and Teachers*, ed. Richard Watt (Edward Elgar Publishing, 2014), 51.

¹²⁴ Geiger, Frosio, and Bulayenko, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects,” 4.

¹²⁵ Geiger, Frosio, and Bulayenko, 4.

¹²⁶ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 700.

¹²⁷ Carroll, “Copyright and the Progress of Science: Why Text and Data Mining Is Lawful,” 895; Flynn et al., “Implementing User Rights for Research in the Field of Artificial Intelligence,” 4.

¹²⁸ Dusollier, “The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition,” 985.

for an exception to accommodate TDM. Therefore, under the new CDSM Directive, TDM can be done in certain circumstances as envisioned in Articles 3 and 4.

TDM holds significant potential as a scientific research tool for analyzing vast amounts of text and data, offering valuable insights in scientific research.¹²⁹ The idea of “*standing on the shoulders of giants*” emphasizes how academic research builds on the work of others, relying on data shared by fellow researchers, publishers, and institutions.¹³⁰ With the majority of scientific journals now accessible online and an ever-growing number of scientific articles being published annually, the potential for TDM is vast.¹³¹ Studies indicated a gradual increase in global TDM publications, with the EU contributing approximately 28.2% in 2016, before the CDSM Directive.¹³² The difficulty in precisely measuring TDM levels, legal uncertainties stemming from existing copyright rules, and various obstacles like lack of awareness and skills hindered widespread adoption.¹³³ Researchers acknowledged TDM’s potential but highlighted the need for clearer legal frameworks and increased awareness to overcome the barriers and fully leverage TDM for advancing scientific research in the EU.¹³⁴

From the “research” point of view, EU Commission described significant obstacles to its growth of the TDM in the EU.¹³⁵ Before the CDSM Directive, uncertainties in copyright rules regarding TDM coverage by exceptions or the need for

¹²⁹ World Intellectual Property Organization, *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)*, 55.

¹³⁰ Christian Handke, Lucie Guibault, and Joan-Josep Vallbb, “Is Europe Falling Behind in Data Mining? Copyright’s Impact on Data Mining in Academic Research,” *SSRN Electronic Journal*, 2015, 2, <https://doi.org/10.2139/ssrn.2608513>.

¹³¹ “Number of Academic Papers Published Per Year – WordsRated,” accessed January 18, 2024, <https://wordsrated.com/number-of-academic-papers-published-per-year/#>; Lutz Bornmann, Robin Haunschild, and Rüdiger Mutz, “Growth Rates of Modern Science: A Latent Piecewise Growth Curve Approach to Model Publication Numbers from Established and New Literature Databases,” *Humanities and Social Sciences Communications* 8, no. 1 (October 7, 2021): 1–15, <https://doi.org/10.1057/s41599-021-00903-w>.

¹³² Sergey Filippov and Paul Hofheinz, “Text and Data Mining for Research and Innovation,” *The Lisbon Council*, no. Issue 20/2016 (2016), https://lisboncouncil.net/wp-content/uploads/2020/08/LISBON_COUNCIL_Text_and_Data_Mining_for_Research_and_Innovation.pdf.

¹³³ European Commission, Impact Assessment on the modernisation of EU copyright rules.

¹³⁴ European Commission.

¹³⁵ European Commission.

specific permissions posed challenges for research organizations.¹³⁶ TDM research faced obstacles from rent-seeking practices and contractual and technical measures.¹³⁷ Providers' conditions complicated access, and technical security measures impeded TDM. Contractual limitations applied even to materials without IP protection.¹³⁸ The diversity in licensing practices added to the challenge. Subscriptions to scientific publications may or may not include permission for TDM, making it unclear for researchers. The risk of conflicting conditions and policies from different publishers was seen as a problem, especially considering the large amount of material involved in scientific research. This complexity led to increased costs for research organizations, forcing them to navigate through different conditions, potentially requiring renegotiations and extra licenses. A leading UK research university (University College London) has reported significant costs, up to GBP 500,000 per year, to ensure compliance with various licenses for their TDM activities.¹³⁹ This highlighted in the Commission's Impact Assessment report that the urgent need for streamlined solutions to address TDM licensing complexities and reduce the financial burden on research institutions.¹⁴⁰

From the technology companies' point of view, the intervention in TDM expected to benefit them indirectly by improving the legal framework. This improvement is likely to create more market opportunities for technology companies, particularly Small and Medium sized Enterprises (SMEs) and start-ups, to serve as partners or service providers for research organizations involved in TDM activities.¹⁴¹

As a result, - although TDM not necessarily should be considered within the scope of copyright according to some scholars¹⁴²- certain actions within TDM research

¹³⁶ Geiger, Frosio, and Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," 13.

¹³⁷ Geiger, Frosio, and Bulayenko, 13.

¹³⁸ Geiger, Frosio, and Bulayenko, 13.

¹³⁹ European Commission, Impact Assessment on the modernisation of EU copyright rules.

¹⁴⁰ European Commission.

¹⁴¹ European Commission.

¹⁴² Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 689; Flynn et al., "Implementing User Rights for Research in the Field of Artificial Intelligence"; Sag, "The New Legal Landscape for Text Mining and Machine Learning"; Carys J Craig, "Globalizing User Rights-Talk: On Copyright Limits and Rhetorical Risks," *American University International Law Review* Vol 33, no. No. 1 (December 2, 2017), <https://ssrn.com/abstract=3109777>.

could potentially violate intellectual property rights. To be more specific, the rights granted by Directive 2001/29/EC (InfoSoc Directive) and Directive 1996/9/EC (Database Directive) may be affected by TDM activities which is the legal framework concerned TDM before the CDSM Directive.¹⁴³ Activities involving reproductions, such as creating copies of protected works in the TDM process, which might trigger protection described in Infosoc Directive.¹⁴⁴ Whereas the mining stage, depending on the software and extraction used, may also infringe on the right of reproduction as stated in the Database Directive.¹⁴⁵ Therefore, the sui generis database right in the Database Directive and InfoSoc Directive will be explored further to enhance understanding of the TDM context.

2.1- Sui Generis Database Right

In the 1990s, the EU introduced the Database Directive (Directive 96/9/EC) to encourage investment in database creation and enhance the information market.¹⁴⁶ This directive aimed to harmonize database protection laws, promote investment in database production for global competitiveness, and ensure a fair balance between users and makers of databases.¹⁴⁷ This initiative originated from previous national rulings that denied copyright protection to collections lacking originality, such as dictionaries.¹⁴⁸ With the rise of electronic databases driven by software, which were vulnerable to easy replication, the need for additional protection for databases became

¹⁴³ Geiger, Frosio, and Bulayenko, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects,” 6.

¹⁴⁴ Article 5 “Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society,” 167 OJ L § (2001), <http://data.europa.eu/eli/dir/2001/29/oj/eng>.

¹⁴⁵ Article 9 “Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases,” 077 OJ L § (1996), <http://data.europa.eu/eli/dir/1996/9/oj/eng>.

¹⁴⁶ Recital 3 and 9 of the Directive Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

¹⁴⁷ Kristina Karanikolova et al., *Study in Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases: Final Report* (Publications Office of the European Union, 2018), 2, <https://data.europa.eu/doi/10.2759/04895>.

¹⁴⁸ Estelle Derclaye and Martin Husovec, “Sui Generis Database Protection 2.0: Judicial and Legislative Reforms,” *European Intellectual Property Review (EIPR)* - Forthcoming, November 16, 2021, 1, Available at SSRN: <https://ssrn.com/abstract=3964943> or <http://dx.doi.org/10.2139/ssrn.3964943>.

apparent.¹⁴⁹ In response, the European Commission proposed measures to safeguard databases by focusing on the investment made in their development rather than solely on their originality.

The *sui generis* database right offers legal protection to databases that have undergone significant investment in acquiring, verifying, or presenting their contents to prevent unauthorized extraction or re-use of substantial portions of the database.¹⁵⁰ Apart from copyright safeguarding databases as original creations, the non-original segment of a database can also receive protection through the *sui generis* database right.¹⁵¹ It provides exclusive rights to the maker of a database, granting protection to the investment made in acquiring, verifying, or presenting the contents of the database.¹⁵² More precisely, the *sui generis* right protects against the unauthorized extraction or re-utilization of database contents.¹⁵³ This right lasts 15 years but resets upon modification, essentially having an indefinite duration.¹⁵⁴ Exceptions include non-commercial teaching, scientific research, private use, and public security, administrative procedures.¹⁵⁵ With the recent exceptions that included with the CDSM Directive, TDM for scientific research doesn't infringe *sui generis* database right in any condition and commercial purposes doesn't infringe unless expressly prohibited, as it discussed in detail under chapter three.

In the context of creating an electronically accessible database, protection extends to both its content and structure. The *sui generis* right covers the content, while copyright safeguards the structure. For instance, copyright applies to original databases, such as those found on websites like Airbnb, where the arrangement reflects

¹⁴⁹ Derclaye and Husovec, 1.

¹⁵⁰ Article 7.1 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

¹⁵¹ Christensen, "A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence," 20.

¹⁵² "Database Protection in the EU," Your Europe, accessed December 29, 2023, https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_en.htm.

¹⁵³ "Database Protection in the EU."

¹⁵⁴ Article 10 (3) Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

¹⁵⁵ Article 9 "Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the Legal Protection of Databases," 077 OJ L § (1996), <http://data.europa.eu/eli/dir/1996/9/oj/eng>.

the author's creative selection and organization of information.¹⁵⁶ However, if technical constraints limit creative choices and result in a basic arrangement, such as chronological or alphabetical order, copyright does not apply.¹⁵⁷

European copyright standards judge artistic works based on originality and trademarks based on consumer perception, allowing users to determine their legal protection.¹⁵⁸ However, sui generis database protection differs by prioritizing investment in data collection, verification, or presentation, leading to an unclear and inconsistent threshold for protection. This means that because the threshold for protection is unclear, it may lead to disparate outcomes for similar datasets. This approach raises concerns about competition, as misuse of database rights could stifle innovation and inflate costs in competitive markets.¹⁵⁹ Diverging from copyright, sui generis right does not require formalities, which may create ambiguity for users seeking clarity on its applicability.¹⁶⁰ This lack of transparency contributes to legal uncertainty, encouraging users to adopt a cautious approach by assuming that the database protected under the sui generis right.

Since the acceptance of this unique type of protection, concerns persist regarding the effectiveness of it, with doubts raised by various stakeholders, both database makers and users, including the European Commission itself.¹⁶¹ In this sense, European Commission conducted two evaluations on the Database Directive aiming to assess whether the Directive met its policy objectives and if the sui generis database right hindered competition. First evaluation of the Database Directive's impact since its inception conducted in December 2005.¹⁶² In this study, The Commission found no

¹⁵⁶ Ducato and Strowel, "Limitations to Text and Data Mining and Consumer Empowerment," 653.

¹⁵⁷ Ducato and Strowel, 653.

¹⁵⁸ Derclaye and Husovec, "Sui Generis Database Protection 2.0: Judicial and Legislative Reforms," 2.

¹⁵⁹ Derclaye and Husovec, 2.

¹⁶⁰ Ducato and Strowel, "Limitations to Text and Data Mining and Consumer Empowerment," 657.

¹⁶¹ Alexander Bernier, Christian Busse, and Tania Bubela, "Public Biological Databases and the Sui Generis Database Right," *IIC - International Review of Intellectual Property and Competition Law* 54, no. 9 (October 1, 2023): 1349, <https://doi.org/10.1007/s40319-023-01373-0>.

¹⁶² "DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases" (Brussels: European Commission, December 12, 2005),

evidence to suggest that the database right increased investment or production in the EU.¹⁶³ Instead, it strengthened monopolies among major players.¹⁶⁴ Ambiguous language in the Directive led to conflicting interpretations among national courts and hindered harmonization efforts.¹⁶⁵ Industry members generally supported the database right, but research and academic institutions criticized its limitations.¹⁶⁶

Second evaluation took place in 2018, reinforced these findings, focusing on industry perspectives and suggesting reform recommendations.¹⁶⁷ It found that giving exclusive rights to database creators might stop useful activities because the real value of data comes from combining and analyzing existing sets.¹⁶⁸ Also, there aren't many exceptions to this right, making it hard to use existing databases.¹⁶⁹ There's also confusion about what the right covers, as the application differs through the member states.¹⁷⁰ As a result, despite its aim to boost investment in database creation, the Database Directive has inadvertently hindered data reuse over the years of its implementation.¹⁷¹ In this legal framework, large organizations navigate its complexities better than smaller ones, reinforcing information monopolies and undermining EU policy goals.¹⁷² Moreover, determining database rights imposes high costs on EU organizations, putting them at a disadvantage. Consequently, the report stated that legislative review is necessary to address unintended consequences.

Within this perspective, the European Commission's 2021 Action Plan on Intellectual Property announced a review of the Database Directive under the Data

https://openfuture.eu/wp-content/uploads/2021/12/2019EC-evaluation_report_legal_protection_databases_december_2005_en.pdf.

¹⁶³ “DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases,” 5.

¹⁶⁴ “DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases,” 20.

¹⁶⁵ “DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases,” 11.

¹⁶⁶ “DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases,” 21.

¹⁶⁷ Karanikolova et al., *Study in Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases*.

¹⁶⁸ Karanikolova et al., 26.

¹⁶⁹ Karanikolova et al., 18.

¹⁷⁰ Karanikolova et al., 7–10.

¹⁷¹ Bernier, Busse, and Bubela, “Public Biological Databases and the Sui Generis Database Right,” 1353.

¹⁷² Bernier, Busse, and Bubela, 1353–54.

Act.¹⁷³ The Data Act, adopted in December 2023, clarifies that the sui generis database right does not apply to databases containing data generated or obtained from Internet of Things (IoT) devices.¹⁷⁴ Article 35 and two recitals (Rec. 84 and 63) address the sui generis database right, exempting databases containing IoT-generated data from sui generis database right application to ensure user data access and sharing rights. This empowers users to utilize and exchange data from various IoT sources, “from coffee machines to wind tribunes”, promoting collaboration, innovation, and user autonomy in the digital realm.¹⁷⁵ This marks the first legislative intervention since the sui generis database right inception in 1996. While not a comprehensive review, it promotes competition, user choice, and data sharing.¹⁷⁶

This legislative intervention that promotes competition, user choice, and data sharing, aligns with the aims of the Court of Justice of the European Union (CJEU)’s recent judgment in the case of CV Online Latvia vs Melons.¹⁷⁷ In a recent judgment, the assessment of whether sui generis database rights are infringed represents a “*significant departure from previous case law*.”¹⁷⁸ In this case, CV-Online, a Latvian company, operates a website with a job advertisement database, using meta tags to aid search engine indexing. Melons, another Latvian company, operates a job search engine that scrapes meta tags from CV-Online’s website. The dispute revolves around whether this constitutes infringement of CV-Online’s sui generis database right. The

¹⁷³ “Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions- Making the Most of the EU’s Innovative Potential An Intellectual Property Action Plan to Support the EU’s Recovery and Resilience” (2020), <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020DC0760>.

¹⁷⁴ “Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Fair Access to and Use of Data (Data Act)” (2022), <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2022:68:FIN>.

¹⁷⁵ “Data Act: Council Adopts New Law on Fair Access to and Use of Data,” accessed April 6, 2024, <https://www.consilium.europa.eu/en/press/press-releases/2023/11/27/data-act-council-adopts-new-law-on-fair-access-to-and-use-of-data/>.

¹⁷⁶ Thomas Margoni, Thomas Gils, and Eyüp Kun, “Chapter X of the Data Act and the Sui Generis Database Right,” CiTiP blog, June 14, 2022, <https://www.law.kuleuven.be/citip/blog/chapter-10-of-the-data-act-and-the-sui-generis-database-right/>.

¹⁷⁷ CV Online Latvia v Melons, No. C-762/19, ECLI:EU:C:2021:434. (Court of Justice of the European Union June 3, 2021).

¹⁷⁸ Derclaye and Husovec, “Sui Generis Database Protection 2.0: Judicial and Legislative Reforms,” 6.

CJEU broadens the issue beyond hyperlinks and meta tags, focusing on the infringement test and balancing the rights of database makers with those of users and competitors. The decision is considerably different from previous case law, emphasizing a more balanced approach to protecting database rights while considering the interests of all parties involved.¹⁷⁹ In this case, the Attorney General suggested that assessing a breach of the sui generis right should consider more than just content extraction, it should also evaluate the harm to the fundamental purpose of the right: protecting the substantial investment of the database producer.¹⁸⁰ In the end, the CJEU ruled that Melons' operation of a specialized search engine, which scraped meta tags from CV-Online's website, constituted infringement of CV-Online's sui generis database right. The CJEU's decision highlights the importance of considering the potential risk to the substantial investment of the database maker while also recognizing the contribution of compilers like Melons to the creation and distribution of innovative products and services in the information sector. Ultimately, the Court left it to the referring Latvian court to apply these criteria to the specific case at hand.

This case holds substantial significance for the research question of this thesis, as it calls into question the efficacy of the sui generis database right, an intellectual property tool intended to foster innovation, after nearly two decades of its implementation. Furthermore, it sheds light on the intricate relationship between the sui generis right and the exceptions made for TDM. Considering that this decision is dates after the TDM exceptions accepted, the Court's reassessment of sui generis database right infringement could have far-reaching implications for these exceptions. For example, it is presumed that it could make the TDM exceptions unnecessary or even broader.¹⁸¹ Hence TDM typically doesn't harm the rights of sui generis database holders, theoretical exceptions might not be as effective as assumed in practice. Nevertheless, it remains uncertain whether the Court will adopt this stance, which has the potential to undermine the position of database producers. If that would be the case,

¹⁷⁹ Derclaye and Husovec, 6.

¹⁸⁰ Tatiana Synodinou, "Search Engines and Databases in the Search for a Balance: The AG's Opinion in the 'CV-Online Latvia' Case," Kluwer Copyright Blog, February 22, 2021, <https://copyrightblog.kluweriplaw.com/2021/02/22/search-engines-and-databases-in-the-search-for-a-balance-the-ags-opinion-in-the-cv-online-latvia-case/>.

¹⁸¹ Derclaye and Husovec, "Sui Generis Database Protection 2.0: Judicial and Legislative Reforms," 11.

producers would need to prove significant investment and how the activity endangers their ability to recover it, especially by risking revenue from database use.¹⁸²

After all these developments on the issue, it can be said that the Database Directive introduces complexities for AI development, particularly regarding the repeated and systematic extraction of insubstantial parts from protected databases.¹⁸³ This prohibition applies across the stages of data mining: during preparation, extraction, and recombination.¹⁸⁴ These activities may require prior authorization from the database maker to safeguard their investments. Before the TDM exceptions in the CDSM, the primary challenge associated with TDM lied in the expansive interpretation of reproduction (in the context of copyright) and extraction (related to the database right). The essential steps in the TDM process, including copying, transferring, and technical adaptation of content, might be eligible for copyright or be protected under the database right.¹⁸⁵ The right to reproduce belongs to the copyright owner, while the right of extraction, even on a temporary basis like visualization on a computer screen, is held by the creator of the database.¹⁸⁶

For instance, if an AI developer wants to make the AI system create songs, the author's rights for the text, the rights of the one who recorded the sounds, and the rights of the database owner where the songs are stored must take into consideration.¹⁸⁷ This shows how TDM adds complexity when used in AI applications. Thus, it can be said that the need for authorization from the rightful owner, or alternatively, reliance on a copyright exception, exposed the limitations in the legal framework for TDM activities, hindering the progress of AI development.

¹⁸² Synodinou, "Search Engines and Databases in the Search for a Balance."

¹⁸³ Romain Meys, "Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?," *GRUR International* 69, no. 5 (May 1, 2020): 461, <https://doi.org/10.1093/grurint/ikaa046>.

¹⁸⁴ Meys, 461.

¹⁸⁵ Geiger, Frosio, and Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," 7.

¹⁸⁶ Ducato and Strowel, "Limitations to Text and Data Mining and Consumer Empowerment," 670.

¹⁸⁷ Christensen, "A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence," 21.

2.2- Information Society Directive

Directive 2001/29/EC, known as the “InfoSoc Directive”, aimed to update copyright rules for digital networks and enact the 1996 WIPO Internet Treaties.¹⁸⁸ By harmonizing critical exclusive rights, such as the right to prevent unauthorized copying (reproduction) and the right to prevent unauthorized online dissemination (making available), the directive sought to establish a comprehensive legal framework. Additionally, it addressed exceptions to these exclusive rights.¹⁸⁹ However, challenges emerged due to the InfoSoc Directive’s broad definition of the reproduction right, encompassing various forms and means of reproduction.¹⁹⁰ This expansive definition led to potential conflicts with copyright law, particularly when TDM involved copyrighted content. Additionally, as explained above, the sui generis database right became a factor too in this context.

Directive 2001/29/EC Article 5(1) offers an exemption for temporary acts of reproduction which is essential to technological processes. The provision aims to facilitate lawful transmission within networks or the legitimate use of a work and to qualify, these acts must be transient, incidental, lack independent economic significance, and play a crucial role in the technological process.¹⁹¹

Despite Art. 5(1) of the InfoSoc Directive covering temporary reproductions under certain circumstances, and from the perspective of TDM activities and computational uses showing potential, its’ conditions are occasionally perceived as restrictive and unclear.¹⁹² Consequently, it described as falling short of becoming a comprehensive framework for various activities dependent on text and data analytics.¹⁹³ While some scholars proposed an interpretive approach to limit the

¹⁸⁸ European Commission, Impact Assessment on the modernisation of EU copyright rules.

¹⁸⁹ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

¹⁹⁰ Article 2, Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

¹⁹¹ Ducato and Strowel, “Limitations to Text and Data Mining and Consumer Empowerment,” 660.

¹⁹² Ducato and Strowel, 662.

¹⁹³ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 691–92.

reproduction right's scope, excluding non-exploitative activities like mining, the EU legislature chose to introduce an explicit TDM exception to ensure legal clarity.¹⁹⁴

2.3- Infopaq Cases

Infopaq cases considered as a landmark judgement of the CJEU, since they significantly shaped and interpreted EU copyright law.¹⁹⁵ These cases involved a media monitoring agency, Infopaq, reproducing small parts of news articles for subscribers.¹⁹⁶ They present an importance in the context of TDM because the Court redefined the originality requirement for a “work”, which in this case it was “11 consecutive words”¹⁹⁷ and interpreted the exception for temporary reproduction which is regulated in Art. 5(1) of the Infosoc Directive.¹⁹⁸

The CJEU, through Infopaq cases I¹⁹⁹ and II²⁰⁰, clarified that under certain conditions, temporary reproductions occurring within “data capture” processes may qualify for the exemption outlined in Article 5(1).²⁰¹ These conditions; the temporary

¹⁹⁴ Hugenholtz, “The New Copyright Directive.”

¹⁹⁵ Estelle Derclaye, “The Court of Justice Copyright Case Law: Quo Vadis?” 36, no. 11 (November 1, 2014), <https://nottingham-repository.worktribe.com/index.php/output/14319944/the-court-of-justice-copyright-case-law-quo-vadis>; Mireille van Eechoud, “Along the Road to Uniformity – Diverse Readings of the Court of Justice Judgments on Copyright Work,” 2012, 60; Eleonora Rosati, “Copyright at the CJEU: Back to the Start (of Copyright Protection),” in *Developments and Directions in Intellectual Property Law*, ed. Hayleigh Bosher and Eleonora Rosati, 1st ed. (Oxford University Press Oxford, 2023), 10, <https://doi.org/10.1093/oso/9780192864475.003.0013>.

¹⁹⁶ Derclaye, “The Court of Justice Copyright Case Law.”

¹⁹⁷ para 60 “Infopaq International A/S v. Danske Dagblades Forening (Case C-5/08), [2012] BUS LR 102 – Center for Study and Research in Intellectual Property Rights [CSRIPR], NUSRL,” accessed January 18, 2024, <https://csriprnusrl.wordpress.com/2022/04/13/infopaq-international-a-s-v-danske-dagblades-forening-case-c-5-08-2012-bus-lr-102/>.

¹⁹⁸ para 63 “Infopaq International A/S v. Danske Dagblades Forening (Case C-5/08), [2012] BUS LR 102 – Center for Study and Research in Intellectual Property Rights [CSRIPR], NUSRL”; This decision also carries importance for press publishers right, see: Cahit Suluk, “Avrupa Birliği Hukukunda Basın Yayıncılara Tanınan Bağlantılı Haklar,” *İstanbul Medeniyet Üniversitesi Hukuk Fakültesi Dergisi* 7, no. 12 (2022): 187.

¹⁹⁹ Court of Justice of the European Union, Opinion of Advocate General Trstenjak delivered on 12 February 2009. Infopaq International A/S v Danske Dagblades Forening. Reference for a preliminary ruling: Højesteret - Denmark. Case C-5/08. (ECJ 2009).

²⁰⁰ Court of Justice of the European Union, Order of the Court (Third Chamber), 17 January 2012. Infopaq International A/S v Danske Dagblades Forening. Case C-302/10. (ECJ 2012).

²⁰¹ Court of Justice of the European Union, Order of the Court (Third Chamber), 17 January 2012. Infopaq International A/S v Danske Dagblades Forening. Case C-302/10.

nature, being an inherent component of a technological procedure, serving the singular objective of facilitating lawful use, lack of independent economic significance, and no modification to the original work, were meticulously outlined by the court.²⁰²

In the Infopaq cases, the CJEU examined a media monitoring service's activity, involving the assembling, extraction, indexing, and printing of newspaper articles and keywords.²⁰³ The court identified five phases in the data capture process, and it ruled that the exception of Article 5(1) covers activities up to the fourth phase but not the printing in the fifth phase, as printing is considered a permanent act of reproduction.²⁰⁴

Within this background, it is suggested that Article 5(1) might facilitate the technological advancement since the data capture process explained in the case shows resemblance to the TDM process.²⁰⁵ However, considering its conditions that are limited to temporary reproductions, and lacking clarity for confident operation in diverse human activities using text and data analytics make it not an ideal legal framework for TDM.²⁰⁶ In this sense, the chosen path by EU will be explained in Chapter 3, whereas what might be more ideal will be further discussed in Conclusion after making some comparisons with other countries.

²⁰² *Opinion of Advocate General Trstenjak delivered on 12 February 2009. Infopaq International A/S v Danske Dagblades Forening. Reference for a preliminary ruling.*

²⁰³ "Infopaq International A/S v. Danske Dagblades Forening (Case C-5/08), [2012] BUS LR 102 – Center for Study and Research in Intellectual Property Rights [CSRIPR], NUSRL."

²⁰⁴ Court of Justice of the European Union, Order of the Court (Third Chamber), 17 January 2012. Infopaq International A/S v Danske Dagblades Forening. Case C-302/10.

²⁰⁵ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 692.

²⁰⁶ Margoni and Kretschmer, 694.

Chapter Three: The TDM exceptions

CDSM Directive represents a comprehensive effort to modernize EU copyright law, aligning it with the demands of the digital age. Comprising 32 Articles and 86 Recitals, the directive addresses various facets of copyright, including exceptions and limitations (Arts.3-7), out-of-commerce works, and licensing practices (Arts. 8-12). The broadness of its coverage highlights the Directive's ambitious goal of adapting copyright regulations to the evolving landscape of the digital single market.²⁰⁷ Articles 3 and 4 of the Directive, primary subject of this thesis, introduce two mandatory TDM exceptions. Article 3 recognizes the inherent challenges tied to TDM, where the reproduction of copyrighted materials and extraction of content from protected databases necessitate the consent of relevant rights holders, absent any applicable exception.²⁰⁸ Acknowledging the importance of fostering scientific research while respecting intellectual property rights, this article establishes a pivotal exception to both copyright and the sui generis right.²⁰⁹ This exception specifically applies to reproductions and extractions carried out by research organizations and cultural heritage institutions during their TDM activities. On the other hand, Article 4 fosters innovation within the private sector by introducing a second exception to both copyright and the sui generis database right. This provision authorizes the reproduction and extraction of lawfully accessible content for TDM purposes.²¹⁰ By doing so, this article provides legal clarity for actions that may not adhere to the conditions of the temporary and transient copy exception outlined in Article 5(1) of the InfoSoc Directive.²¹¹

Understanding TDM rights in the EU requires clarity on how they apply in both scientific and commercial contexts. While the necessity of ensuring broad TDM rights in the scientific domain has been evident, the EU's position on TDM in the commercial sector has been less certain.²¹² This uncertainty was evident in the initial draft of the

²⁰⁷ Margoni and Kretschmer, 685.

²⁰⁸ Meys, "Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market," 465.

²⁰⁹ Meys, 465.

²¹⁰ Meys, 466.

²¹¹ Quintais, "Generative AI, Copyright and the AI Act."

²¹² Hugenholtz, "The New Copyright Directive."

CDSM Directive by the European Commission, where a mandatory TDM exception was proposed exclusively for non-commercial research institutions and cultural heritage organizations.²¹³ After the discussion process, the CDSM Directive incorporated two obligatory TDM provisions, namely Articles 3 and 4, each with distinct implications and varying strengths. This intricate evolution of TDM provisions highlights the approach taken by the EU in reconciling the scientific and commercial dimensions of TDM.

Article 25 of CDSM states that ‘*Member States may adopt or maintain in force broader provisions, compatible with the exceptions and limitations provided for in the Database and InfoSoc Directives, for uses or fields covered by the exceptions or limitations provided for in this Directive*’.²¹⁴ With this provision, it is possible that Member States might prefer to implement a wider TDM exception, usually referring to the article 5(3)(a) of the Infosoc Directive. Even though this article allows broader TDM exception, it depends on each Member States’ acts and doesn’t act as a comprehensive rule for all EU.²¹⁵ Still, France, Estonia and Germany are some of the member states that seized this opportunity to give TDM exceptions more flexibility.²¹⁶

²¹³ “The Copyright Directive Is Failing (Open Letter to Members of the European Parliament and the Council of the European Union),” accessed December 10, 2023, https://www.create.ac.uk/wp-content/uploads/2018/04/OpenLetter_EU_Copyright_Research_Centres_26_04_2018.pdf.

²¹⁴ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²¹⁵ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 691.

²¹⁶ Christensen, “A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence,” 26; Maria Bottis et al., “Text and Data Mining in the EU ‘Acquis Communautaire’ Tinkering with TDM & Digital Legal Deposit,” *Erasmus Law Review* 12, no. 2 (2019): 190–208, <https://doi.org/10.5553/ELR.000140>.

PROVISION 	Article 3	Article 4
SCOPE & CONDITIONS 		
Acts	Text and Data Mining	Reproductions and Extractions
Rights	Database Dir.: 5(a) and 7(1) (The rights under Article 5(a) and Article 7(1) pertain to the rights of makers of databases to control unauthorized extraction and/or re-utilization of the contents of their databases.) InfoSoc Dir.: 2 (The rights under Article 2 relate to the exclusive rights granted to authors of literary and artistic works, including the right of reproduction, distribution, communication to the public, and adaptation.) CDSM Dir.: 15(1) (The rights under Article 15(1) likely concern the rights and obligations related to certain online platforms and their activities in the digital single market, ensuring fair and transparent trading practices)	Database Dir.: 5(a) and 7(1) InfoSoc Dir.: 2 Software Dir.: 4(1)(a) and (b) (The rights under Article 4(1)(a) and (b) involve the exclusive rights of software authors, such as the right to reproduce the software and the right to translate, adapt, and alter the software.) CDSM Dir.: 15(1)
Purpose of Use	Scientific research	Text and data mining
Beneficiaries	Research organizations; Cultural heritage institutions	All (unrestricted)
Lawful access	Yes	Yes
Other Conditions for Beneficiary	Works must be stored with “appropriate level of security”	-
Contractual Derogation	No	Yes
Other Reservations or Exclusions	No	Yes: via “machine-readable means in the case of content made publicly available online”
Connection to other exceptions in acquis	InfoSoc Dir.: 5(3)(a) (This article outlines an exception to copyright, allowing for the reproduction of a work by libraries, educational establishments, museums, and archives for the purpose of preservation of cultural heritage.)	InfoSoc Dir.: Art. 5(1) (This article provides for a general exception to copyright, allowing member states to introduce limitations to copyright for certain specific purposes such as criticism, review, and reporting current events, as long as they are carried out in accordance with fair practice and do not conflict with the normal exploitation of the work.)

Table 1: TDM Exceptions in the CDSM Directive ²¹⁷

3.1- Article 3 – Scientific Research Exception

The primary objective of Article 3 is to establish a compulsory exception within the framework of EU copyright law. This exception applies specifically to acts of reproduction (for copyright-protected material) and extraction (for the sui generis database right) conducted by research organizations and cultural heritage institutions, collectively referred to as research and cultural organizations.²¹⁸ The purpose of these acts is to facilitate TDM for scientific research purposes. What might be “scientific research” is briefly touched upon in Recital 12, stating that both natural and human sciences are covered with this definition.²¹⁹

It must be noted that Article 3 of the law allows only certain organizations, research institutions and cultural heritage entities such as libraries and museums, to reproduce and extract content for their activities. By doing so, it excludes some entities like public broadcasters and commercial research institutes from these allowances, though they might have alternatives under Article 4. Additionally, Article 3(2) not only permits mining activities but also allows the safe storage of mined works for scientific research, which is considered as a crucial need since empirical scientific research typically requires research data to remain accessible for verification purposes.²²⁰ Article 3 only allows mining for works to these organizations provided they have “lawful access” to the relevant works or subject matter. Referred to as the “pre-existing” condition, this means that research organizations and cultural heritage institutions must already have legal access to the resource in question.²²¹ The directive defines lawful access broadly, encompassing open access policies, contractual

²¹⁷ João Pedro Quintais, “The New Copyright in the Digital Single Market Directive: A Critical Look,” *SSRN Electronic Journal*, 2019, 10, <https://doi.org/10.2139/ssrn.3424770>.

²¹⁸ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²¹⁹ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²²⁰ Hugenholtz, “The New Copyright Directive”; Carroll, “Copyright and the Progress of Science: Why Text and Data Mining Is Lawful,” 908.

²²¹ Ducato and Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 6.

arrangements, subscriptions, and other lawful means, including freely available online content.²²² Importantly, this doesn't mean copyright holders can ban mining through contracts; Article 7 makes any such provision unenforceable. Also, the choice to "opt out" of the mining exemption is only available for non-research uses covered by Article 4.

3.2- Article 4 - General Exception

In parallel to Article 3, Article 4 closely mirrors the provisions of Article 3, with a notable distinction: It extends accessibility to any beneficiary for any use. It allows copy and extract content for TDM purposes, keeping these copies for the same use.²²³ This covers various purposes, including commercial research, journalism, data analysis, and AI.²²⁴ However, Article 4 introduces a significant difference - rightsholders retain the option to explicitly reserve this exception, allowing for its potential override through mechanisms such as "opt-out" or "contract-out". In other terms, it is only possible when copyright holders must not have explicitly prohibited this use through contracts, declarations, machine-readable means, or terms and conditions. This means that copyright owners can still give permission for TDM activities and charge fees for data usage.²²⁵ Therefore, if copyright holders haven't clearly "*license the uses of their works or other subject matter they want to keep their rights*" (in a way like machine-readable methods), as suggested in Recital 18, then a specific part of Article 4 (3) comes into play.²²⁶

Article 4(3) of the Directive outlines a condition for the retention of the right to prohibit TDM activities. This condition is that copyright holders must not have explicitly reserved their rights in a way that is considered appropriate. The appropriateness, as specified in Recital 18, primarily involves using machine-readable means like metadata or website terms and conditions when content is made publicly

²²² Ducato and Strowel, 6.

²²³ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²²⁴ Dusollier, "The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition," 986.

²²⁵ Dusollier, 987.

²²⁶ Hugenholtz, "The New Copyright Directive."

available online.²²⁷ The emphasis is on machine-readable means, particularly for content available online, to provide clear and accessible information about the rights associated with TDM activities.²²⁸ This ensures that users engaging in TDM are aware of any restrictions or permissions set by the copyright holders. In other situations, copyright holders can reserve their rights through alternative means such as contractual agreements or a one-sided declaration.

²²⁷ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²²⁸ Hugenholtz, “The New Copyright Directive.”

Chapter Four: Assessment of the Articles

International principle of copyright protects unique creative works, while ideas, concepts, fundamental principles, methods, factual information, and data in their raw form do not receive protection. Bearing in mind of this nature of the copyright, there is an opinion in the literature points out that TDM exception would be meaningless in an ideal copyright framework because the extraction of factual information from copyrighted material is beyond the reach of copyright law.²²⁹

Despite the argument that this copying is technical and not an act of exploitation, the EU chose to make an exception to accommodate TDM. Article 3 of the Directive exclusively caters to research and cultural organizations, outlining a limited scope for TDM exceptions. Conversely, Article 4 of the Directive takes an opt-out approach, potentially impeding commercial TDM activities. The current framework concentrates on restricting access to these tools to specific entities, mainly research-oriented, potentially limiting the broader potential of TDM. A more inclusive approach could have been pursued by expanding the TDM exception to include individuals and entities beyond the research sector, such as start-ups and journalists.²³⁰ Such an expansion would democratize access to TDM tools, recognizing the diverse contributors to knowledge creation and fostering a more inclusive research ecosystem extending beyond traditional organizations.²³¹ Hence, Articles 3 and 4 of the CDSM Directive are perceived as problematic from five different perspectives, encompassing the definition of TDM, the beneficiaries, rights, the capacity for technological adaptation, and the availability of primary resources.²³²

²²⁹ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 689; See also: Flynn et al., “Implementing User Rights for Research in the Field of Artificial Intelligence”; Sag, “The New Legal Landscape for Text Mining and Machine Learning”; Craig, “Globalizing User Rights-Talk: On Copyright Limits and Rhetorical Risks”; Lemley and Casey, “Fair Learning.”

²³⁰ Geiger, Frosio, and Bulayenko, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects,” 4.

²³¹ Geiger, Frosio, and Bulayenko, 4.

²³² Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 687–88.

4.1- Challenges with TDM Definitions and Beneficiaries

The definition of TDM in the CDSM Directive, outlined in Article 2 as “*any automated analytical technique aiming to analyze text and data in digital form to generate information such as patterns, trends and correlations*” and further enhanced in Recital 8 of the Directive as, “*the automated computational analysis of information in digital form, such as text, sounds, images or data enabled by new technologies*” considered potentially problematic. Concerns raised about the definition states that encompassing a wide range of automated analytical techniques, may be too extensive.²³³ Specifically, it is argued that the definition essentially describes a common task performed by AI-based machine learning systems. Consequently, the significant impact of the Directive’s language on the regulation of AI and ML practices, both currently recognized and those emerging in the future, as long as they rely on data analysis. That is why, starting with the definitions of TDM, exceptions too seemed as “*conceptually wrong, theoretically flawed and normatively unambitious.*”²³⁴

On the other hand, excluding unaffiliated individuals and researchers from the TDM exception in the CDSM Directive contradicts the principle that the “*the right to read is the right to mine*”, which means that extracting information for analytical purposes does not inherently infringe upon copyright or related rights.²³⁵ So, this limitation may impede knowledge democratization, hinder opportunities for start-ups, and hamper potential contributions in various fields, including journalism.²³⁶ The consequence is a setback in addressing contemporary challenges, such as combating fake news.²³⁷

²³³ Margoni and Kretschmer, 688.

²³⁴ Margoni and Kretschmer, 686.

²³⁵ Peter Murray-Rust, “The Right to Read Is the Right to Mine,” Open Knowledge Foundation blog, June 1, 2012, <https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/>.

²³⁶ Geiger, Frosio, and Bulayenko, “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects,” 21.

²³⁷ Erik Kleinsmith, “Combating Fake News with Data Mining,” *Edge* (blog), July 22, 2019, <https://amuedge.com/combating-fake-news-with-data-mining/>.

4.2 - TDM's Role in AI and Copyright Law

In this sense, this explicit authorization highlights a critical aspect of the TDM landscape; the intricate relationship between TDM and access to content, particularly within the context of AI. While regulating a widely used technology like AI may extend beyond the original scope of copyright law, it is crucial to recognize the fundamental role of data in AI systems and their resulting products.²³⁸ Granting property rights in data or imposing restrictions on the use of unprotected factual information will inevitably impact the evolution of this technology.²³⁹ As mentioned in previous chapters, TDM plays a pivotal role as a catalyst for ML and AI advancements. The success of AI innovation, especially in ML, heavily relies on the abundance and quality of available data to train algorithms effectively.²⁴⁰ Which means inadequate, incomplete, or non-representative input data can lead to below mediocre and unreliable outputs.²⁴¹ Therefore, good algorithms need good training data. TDM also holds the potential to address issues of algorithmic bias and discrimination, as evidenced by instances where predictive algorithms discriminated against minorities due to biased training datasets.²⁴² TDM is particularly crucial in fields like medicine, as the unavailability or limited access to diverse datasets can worsen health inequalities.²⁴³ However, the concept of lawful access, primarily enforced through contracts, poses a potential obstacle to achieving such ambitious and necessary goals, particularly in the field of AI. ²⁴⁴ In essence, by formulating the regulations governing access to a particular technology and by distributing ownership of its essential components, would mean that molding that technology and its societal

²³⁸ Senftleben et al., “Ensuring the Visibility and Accessibility of European Creative Content on the World Market,” 71.

²³⁹ Senftleben et al., 71.

²⁴⁰ Ducato and Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 20.

²⁴¹ Abdulaziz Aldoseri, Khalifa N. Al-Khalifa, and Abdel Magid Hamouda, “Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges,” *Applied Sciences* 13, no. 12 (January 2023): 6, <https://doi.org/10.3390/app13127082>.

²⁴² Eli M. Cahan et al., “Putting the Data before the Algorithm in Big Data Addressing Personalized Healthcare,” *Npj Digital Medicine* 2, no. 1 (August 19, 2019): 1–6, <https://doi.org/10.1038/s41746-019-0157-2>.

²⁴³ Mirja Mittermaier, Mariam M. Raza, and Joseph C. Kvedar, “Bias in AI-Based Models for Medical Applications: Challenges and Mitigation Strategies,” *Npj Digital Medicine* 6, no. 1 (June 14, 2023): 1–3, <https://doi.org/10.1038/s41746-023-00858-z>.

²⁴⁴ Ducato and Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 20.

impact for the years ahead.²⁴⁵ Currently, in the EU, these regulations dictate that in order to develop AI, entities such as firms, governments, citizens, journalists, and anyone other than research and cultural organizations acting for research purposes must obtain explicit authorization from rightsholders.

4.3 - Implications on Copyright Principles

The copyright protection measures to the TDM process challenges its' core purpose of fostering information dissemination, a principle embedded in copyright's historical roots dating back over three centuries.²⁴⁶ Originally crafted to reward creators and promote public learning, copyright traditionally protects expressions, not facts or ideas.²⁴⁷ The inclusion of TDM within copyright boundaries raises concerns about the foundational principles of copyright law. The contradiction becomes evident in the CDSM Directive's treatment of factual information and data ownership in the digital age.²⁴⁸ While such elements appear excluded from copyright protection individually, their incorporation into protected works triggers exclusivity rights.²⁴⁹ This challenge is compounded by the prevalence of digital copies, prompting a unique path in EU copyright law that raises questions about originality standards and the reproduction right's broad interpretation for digital uses.²⁵⁰ Scholars like Margoni and Kretschmer argue for a more transparent and explicit legal framework, emphasizing the need to align copyright principles with the evolving nature of information use in the digital era.²⁵¹

²⁴⁵ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 699.

²⁴⁶ Ducato and Strowel, "Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out," 24; Pamela Samuelson, "The Digital Dilemma: A Perspective on Intellectual Property in the Information Age," n.d., <https://people.ischool.berkeley.edu/~pam/papers/digdilsyn.pdf>; *The Digital Dilemma: Intellectual Property in the Information Age*, accessed June 13, 2024, <https://doi.org/10.17226/9601>.

²⁴⁷ Ducato and Strowel, "Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out," 23.

²⁴⁸ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 690.

²⁴⁹ Margoni and Kretschmer, 699.

²⁵⁰ Margoni and Kretschmer, 701.

²⁵¹ Margoni and Kretschmer, 701.

4.4 - Shortcomings of the CDSM Directive

The articles surrounding TDM also criticized when comparing Article 14 of the same directive. The legislative technique used in Article 14 of the CDSM Directive takes a different, more balanced approach compared to Articles 3 and 4.²⁵² Article 14 explicitly states that digitizing visual art does not establish new rights within the copyright or related rights domain. The comparison is drawn to illustrate that a similar approach could have been taken for Articles 3 and 4, where the lawmaker could have specified that extracting non-protected facts and data from copyrighted works does not constitute a copyright infringement.²⁵³

One of the less acknowledged but highly impactful challenges facing TDM in the AI domain is trade secret protection, because the secrecy of important elements like training data limits TDM's role in AI development and its ability to provide explanations.²⁵⁴ Notably, these challenges remain unaddressed in the CDSM Directive.

In light of the challenges posed by the new Directive and its TDM exceptions, it becomes evident that the intended goal of fostering a more innovation-friendly environment may not be fully realized. The Directive's provisions, rather than providing a clear and unified framework, leave room for divergence among Member States as stated by Quintais *"this regime will probably not lead to simplification and harmonization of the system of exceptions in EU copyright law, as it continues to allow significant cherry-picking by the Member States."*²⁵⁵ This potential lack of harmonization may hinder the streamlined and simplified approach that was envisaged, thus falling short of creating the envisioned innovation-friendly landscape across the EU.

4.5- Assessment of Article 3

Article 3, limits TDM exercises by various senses. Formulation of the article indicates that, firstly it can only be used for research purposes and by research and

²⁵² Margoni and Kretschmer, 690.

²⁵³ Margoni and Kretschmer, 690.

²⁵⁴ Ducato and Strowel, "Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out," 31.

²⁵⁵ Quintais, "The New Copyright in the Digital Single Market Directive," 11.

cultural organizations or private ordering. According to Recital 12, not all research organizations are eligible to conduct TDM, they “*must operate on a not-profit basis or reinvest all its profits into its scientific research or pursue a public interest mission funded by public funds or public contracts.*”²⁵⁶ The Directive’s narrow definition of eligible research organizations for TDM may limit participation, especially for commercial-based research entities reliant on private sector funding.²⁵⁷ So the TDM exception is primarily designed for institutions providing cultural or public services on a non-profit basis, whereas the reality of research often involves collaborations with private entities.²⁵⁸ The CDSM Directive addresses this by allowing private partners to benefit from the TDM exception in the context of public-private partnerships, provided it aligns with the project’s needs and is limited to the project’s scope and duration.²⁵⁹ Therefore this emphasis on non-profit or publicly funded structures could unintentionally exclude innovative contributions, posing challenges to fostering a diverse and dynamic TDM landscape and impacting the collaborative nature of scientific and technological advancements.²⁶⁰ Although the legislators’ intention is to maintain the neutrality and independence of TDM-driven scientific research, the scarcity of public funding may limit the capabilities of research organizations, stifling innovation in the EU.²⁶¹ Moreover, the lack of a clear definition for “scientific research” within the directive raises concerns about possible conflicts with existing licenses and the potential for restrictive interpretations, further complicating the landscape for TDM activities.²⁶²

²⁵⁶ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²⁵⁷ Christensen, “A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence,” 30.

²⁵⁸ Ducato and Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 6.

²⁵⁹ Recital 11 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²⁶⁰ Dermawan, “Text and Data Mining Exceptions in the Development of Generative AI Models,” 9.

²⁶¹ Charlotte Gerrish and Anders Molander Skavlan, “European Copyright Law and the Text and Data Mining Exceptions and Limitations” 2, no. 2 (2019).

²⁶² Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, “Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/Eu,” *SSRN Electronic Journal*, 2019,

From the technical perspective, it covers such broad area that almost every automated method that examines digital information falls within the specific limits outlined in the articles.²⁶³ Examples to this to be able to comprehend the extends would be that any data driven AI forms, traditional ML structures, and more sophisticated deep learning forms falls in its context. This outcome of the Directive leads to an understanding that licensing rights and research purposes are prioritized whereas technological perspective were not thoroughly examined, even though this aspect has the greater effect on the competitiveness on the area.²⁶⁴

The crucial prerequisite of lawful access, stated in Article 3(1), explained in Recital 14, which defines as access to content available through open access policies, contracts, subscriptions, or other lawful means, covering freely available online content.²⁶⁵ Examining Article 3 in connection with Recitals 10 and 17 suggests that, in the presence of lawful access, TDM for research should be permitted without additional burdens.²⁶⁶ Recital 10 raises concerns about the potential exclusion of TDM in licenses over content to which research organizations have lawful access, cautioning against disadvantages.²⁶⁷ Meanwhile, Recital 17 emphasizes that Member States should not provide compensation for rightsholders, indicating that any harm to rightsholders through the TDM exception would be ‘minimal’.²⁶⁸ This aligns with the

<https://doi.org/10.2139/ssrn.3470653>. ,Ducato and Strowel, “Ensuring TDM: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 11

²⁶³ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 686.

²⁶⁴ Margoni and Kretschmer, 688.

²⁶⁵ Recital 14 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²⁶⁶ Ducato and Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 19.

²⁶⁷ Recital 10 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²⁶⁸ Recital 17 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

Directive's intent that, where lawful access exists, TDM for research should ideally be unrestricted, possibly without additional compensation.²⁶⁹

Yet, depending on recitals poses difficulties, given their non-binding nature and Recital 17 could be viewed as optional.²⁷⁰ Publishers may capitalize on this by increasing subscription costs for research organizations universally, potentially undermining the Directive's aim to facilitate TDM for research without unnecessary costs.²⁷¹ As mentioned in previous chapters, the practice of charging additional fees for TDM would not be affordable for many important research projects. Relying only on licensed or public domain content would limit research, affect reproducibility, introduce bias, and discourage researchers.²⁷² Hence, not having a license would eventually be a complete barrier to TDM research.

4.6 - Assessment of Article 4

Article 4 of the EU CDSM Directive, particularly its opt-out mechanism, has faced strong criticism for its potential negative impact on AI advancement in the EU.²⁷³ Whereas it can be said that this exception is designed to facilitate AI development and research, the authority granted to rightsholders to restrict certain uses highlights the nuanced balance the European legislator aims to strike.²⁷⁴ The efficacy of these provisions in managing the risk of data blockage while upholding intellectual property rights is a crucial aspect to observe as AI technology evolves.²⁷⁵ Therefore the focus of concern is Article 4(3), allowing rightsholders to reserve TDM rights. The reservation process, outlined in Recital 18, is subject to conditions, distinguishing between publicly available online content and other types. Recital 18 reiterates lawful access as a prerequisite for exercising the TDM exception under Article 4. This includes content made available to the public online, provided rightsholders have not

²⁶⁹ Ducato and Strowel, "Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out," 19.

²⁷⁰ Ducato and Strowel, 19.

²⁷¹ Geiger, Frosio, and Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," 22.

²⁷² Fiil-Flynn et al., "Legal Reform to Enhance Global Text and Data Mining Research," 952.

²⁷³ Geiger, Frosio, and Bulayenko, "Text and Data Mining," 36.

²⁷⁴ Hacker, "A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act," 278.

²⁷⁵ Hacker, 277.

reserved rights to prevent TDM.²⁷⁶ However, challenges arise when websites implement exclusion protocols or terms and conditions that impede crawling, making it challenging to verify reservation of rights.²⁷⁷ Without machine-readable licenses, users performing TDM may need to manually check the terms of each website, an impractical scenario given the automated and diverse nature of TDM activities.²⁷⁸

Article 4 is seen as a problem because it strongly focuses on licensing, favoring private rules over public policies.²⁷⁹ This makes it a big challenge for GenAI developers who need a lot of data, possibly slowing down progress in the field. That is why the ‘opt-out’ nature of this article may go beyond a simple exception of TDM for copyright purposes.²⁸⁰ From the AI and ML perspective, it can be stated that this provision allows copyright holders to opt out of having their works used in training datasets for AI models. Some commentators view this as a positive development, believing it could empower rights holders, potentially leading to licensing agreements and compensation from AI providers.²⁸¹ Conversely, others are critical, expressing concerns that it could concentrate power in the hands of large companies would lead to problems like disadvantaged SMEs or exploitation of creators. For instance, Margoni and Kretschmer argues that the clauses within the CDSM Directive involuntarily promote the creation of biased AI systems.²⁸² This occurs because of the economic incentives provided by the pricing and conditions ensuring the availability

²⁷⁶ Recital 18 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

²⁷⁷ Ducato and Strowel, “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,” 19.

²⁷⁸ Ducato and Strowel, 19.

²⁷⁹ Kollár, “Mind If I Mine? A Study on the Justification and Sufficiency of Text and Data Mining Exceptions in the European Union,” 19.

²⁸⁰ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 688–89.

²⁸¹ Paul Keller, “Protecting Creatives or Impeding Progress? Machine Learning and the EU Copyright Framework,” Kluwer Copyright Blog, February 20, 2023, <https://copyrightblog.kluweriplaw.com/2023/02/20/protecting-creatives-or-impeding-progress-machine-learning-and-the-eu-copyright-framework/>; “Policy Paper #15 on Using Copyrighted Works for Teaching the Machine,” *COMMUNIA Association* (blog), accessed January 15, 2024, <https://communia-association.org/policy-paper/policy-paper-15-on-using-copyrighted-works-for-teaching-the-machine/>.

²⁸² Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 700; See also: Lemley and Casey, “Fair Learning,” 127.

of training data. Hence, developers may find it economically advantageous to bypass licensing requirements by training algorithms on outdated, less accurate, and biased data, or by acquiring AI models trained on unverifiable data.²⁸³ Same way, letting big companies train AI models on public works might be a problem from the creative aspect as well. It is argued that might lead to big companies having more power and exploiting artists.²⁸⁴ Since creative job markets are already controlled by a few big players, these companies could force artists into contracts that make them give up important rights for less pay.²⁸⁵ In the end, this could mean more power for big companies and less control and pay for artists. This situation proves that this Directive, which initially aims a “digital single market”, is ironically strengthens the disadvantages of SMEs against big tech companies.²⁸⁶ Because the aim of creating unified market is problematic given that requirement of licenses scattered through member states and from Article 4 point of view even more through the right holders. In this conditions, big players have more potential to collect as more licenses as possible, which results continuity of their dominance.²⁸⁷

The Directive regulates AI from the point of property rights because its’ building blocks, “the data”, becomes under the protection of copyright. That’s why this Directive might be representing a potentially extensive legislative advancement with the potential to significantly influence the interaction between legal principles and technology, subsequently, the future of the EU’s legal framework.²⁸⁸ In this sense, investigative journalism faces a similar problem, as journalists often use documents from whistleblowers lacking appropriate authorization, and these documents can be

²⁸³ Hacker, “A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act,” 280.

²⁸⁴ Quintais, “Generative AI, Copyright and the AI Act.”

²⁸⁵ Cory Doctorow and Katharine Trendacosta, “AI Art Generators and the Online Image Market,” Electronic Frontier Foundation, April 3, 2023, <https://www EFF.org/deeplinks/2023/04/ai-art-generators-and-online-image-market>.

²⁸⁶ Pamela Samuelson, “Questioning a New Intellectual Property Right for Press Publishers – Communications of the ACM,” March 1, 2019, <https://cacm.acm.org/opinion/questioning-a-new-intellectual-property-right-for-press-publishers/>; Suluk, “Avrupa Birliği Hukukunda Basın Yayıncılara Tanınan Bağlantılı Haklar,” 192.

²⁸⁷ Samuelson, “Questioning a New Intellectual Property Right for Press Publishers – Communications of the ACM”; Suluk, “Avrupa Birliği Hukukunda Basın Yayıncılara Tanınan Bağlantılı Haklar,” 192.

²⁸⁸ Margoni and Kretschmer, “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology,” 688.

configured to block data analysis, making it difficult for them to benefit from the Article 4 exception.²⁸⁹

Against this backdrop, an extraordinary viewpoint regarding the potential obstruction posed by rightsholders exercising their veto power under Article 4(3) of the CDSM Directive suggests exploring an interpretation of the TDM exception that fosters innovation.²⁹⁰ This approach takes into account the delicate balance between the interests of AI developers and rightsholders. Since the potential blockage of access to crucial training data due to copyright restrictions poses a significant risk for impeding innovation in the field of AI, commercial research conducted within companies should also benefit from the Article 3 exception, challenging the current distinction between non-commercial and commercial research.²⁹¹ This proposition, grounded in the principles of equal treatment outlined in primary EU law, specifically Article 20 of the Charter, aims to foster opportunities for AI research within SMEs and other entities, ultimately promoting innovation while respecting copyright and legal frameworks.²⁹² In the end, IP protection mechanisms are designed to encourage invention by ensuring inventors can earn economic returns from their work.²⁹³ This nuanced perspective strives to strike a reasonable balance, ensuring that commercial research also reaps the benefits of the Article 3 exception, contributing to ongoing advances in AI research without compromising copyright principles.

4.6.1- The Technical Implementation of Article 4 Section 3

The opt-out provision (Article 4(3)), makes the EU stand out globally by prioritizing creators' rights to withhold their works from ML training data, distinguishing it from regions like the United States and Japan where regulations lack explicit provisions or have blanket exceptions permitting unrestricted use of copyrighted works for ML training. This provision prioritizes the rights of creators, allowing them to control whether their works are used in ML training data and by

²⁸⁹ Dusollier, "The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition," 987.

²⁹⁰ Hacker, "A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act," 294.

²⁹¹ Hacker, 295.

²⁹² Hacker, 295.

²⁹³ Tuğba Güleş, *Technological Advances in Plant Innovation and the Dynamics of Intellectual Property Law*, Birinci baskı, Seçkin. Hukuk, no: 1195 (Ankara: Seçkin, 2018), 37.

explicitly acknowledging and safeguarding these rights, it also aims to bring clarity to legal aspects for ML researchers and developers.²⁹⁴

From a chronological point of view, it can be stated that TDM exceptions that adopted with CDSM Directive in 2019, before the breakthrough of Generative ML models in 2022²⁹⁵, demonstrates proactive legal framing.²⁹⁶ However, the lack of a widely adopted opt-out mechanism under Article 4 poses challenges; implementing Article 4 for machine-readable opt-outs in ML training faces uncertainty without recognized standards in the wording of it.

In this sense, significant delays which estimated to take up to 1.5 years in ML model training is concerning.²⁹⁷ Simply means that the data chosen for 2022 breakthrough models was selected in early 2021 when awareness of generative ML impact was limited.²⁹⁸ This delay indicates widely accepted practices for opting out copyrighted works from ML training only take effect in trained models after a considerable delay. The delay between opting out and its impact might mean ineffective exercise of opt- out right for creators and rights holders, given that trained ML models find it challenging to “unlearn” information from specific works in their training data.²⁹⁹

In response to these challenges in exercising the right to opt-out, several initiatives have emerged with the goal of empowering creators, rights holders, and developers by establishing effective mechanisms to control the utilization of copyrighted works in ML training datasets. Most known among them; Spawning AI, introduced in September 2022, allows creators to register opt-outs on

²⁹⁴ Paul Keller and Zuzanna Warso, “Defining Best Practices for Opting out of ML Training,” Open Future EU, September 29, 2023, 1, https://openfuture.eu/wp-content/uploads/2023/09/Best-_practices_for_optout_ML_training.pdf.

²⁹⁵ “Generative AI Is Changing Everything. But What’s Left When the Hype Is Gone?,” MIT Technology Review, accessed January 15, 2024, <https://www.technologyreview.com/2022/12/16/1065005/generative-ai-revolution-art/>.

²⁹⁶ Keller and Warso, “Defining Best Practices for Opting out of ML Training,” 5.

²⁹⁷ Keller and Warso, 6.

²⁹⁸ Keller and Warso, 6.

²⁹⁹ Thanveer Shaik et al., “Exploring the Landscape of Machine Unlearning: A Comprehensive Survey and Taxonomy” (arXiv, October 23, 2023), 15, <http://arxiv.org/abs/2305.06360>; Matthew Duffin, “Machine Unlearning: The Critical Art of Teaching AI to Forget,” *VentureBeat* (blog), August 12, 2023, <https://venturebeat.com/ai/machine-unlearning-the-critical-art-of-teaching-ai-to-forget/>.

havebeentrained.com, with a reported 78 million opt-outs by March 2023.³⁰⁰ From a technical point of view, Spawning has a tool that helps AI developers respect creators, choices to opt out of having their content used in AI training.³⁰¹ This tool, called an API, allows this opt-out information to be easily included in the AI development process. Additionally, Spawning introduced “ai.txt”, a feature that lets websites specify whether their content can be used for TDM in a way that machines can understand.³⁰² As demands increased, Spawning launched a new tool, “Kudurru” designed to detect and block web scraping in real-time, providing artists with a proactive means to control and protect their intellectual property from unauthorized use in AI training.³⁰³

On the other hand, Google and Open AI introduced distinct opt-out standards for rights holders, allowing exclusion of works from specific AI models.³⁰⁴ However, these mechanisms, based on robots.txt extensions, misrepresent compliance with Article 4(3) of the CDSM Directive.³⁰⁵ While creators need the right to reserve works for TDM, model-specific opt-outs proved impractical, requiring repetitive actions for each training entity.³⁰⁶

While the approaches to tackle this problem differ among various initiatives, it is suggested that the European Commission, tasked with overseeing the CDSM

³⁰⁰ Mat, “Spawning Opts out 78 Million Artworks from AI Training,” May 30, 2023, <https://spawning.substack.com/p/spawning-opts-out-78-million-artworks>.

³⁰¹ Spawning AI, “Datadiligence Respect Generative AI Opt-Outs in Your ML Training Pipeline,” Spawning AI, accessed January 15, 2024, <https://datadiligence.readthedocs.io/en/latest/readme.html>.

³⁰² Kyle Wiggers, “Spawning Lays out Plans for Letting Creators Opt out of Generative AI Training,” *TechCrunch* (blog), May 3, 2023, <https://techcrunch.com/2023/05/03/spawning-lays-out-its-plans-for-letting-creators-opt-out-of-generative-ai-training/>.

³⁰³ Kate Knibbs, “A New Tool Helps Artists Thwart AI—With a Middle Finger,” *Wired*, accessed January 15, 2024, <https://www.wired.com/story/kudurru-ai-scraping-block-poisoning-spawning/>.

³⁰⁴ Thorin Klosowski, “No Robots(.Txt): How to Ask ChatGPT and Google Bard to Not Use Your Website for Training,” Electronic Frontier Foundation, December 12, 2023, <https://www EFF.org/deeplinks/2023/12/no-robotstxt-how-ask-chatgpt-and-google-bard-not-use-your-website-training>.

³⁰⁵ Paul Keller, “Generative AI and Copyright: Convergence of Opt-Outs?,” Kluwer Copyright Blog, November 23, 2023, <https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/>.

³⁰⁶ Keller.

Directive's implementation, should proactively address this issue.³⁰⁷ A standardized and transparent opt-out provision is crucial to ensure that developers of AI models provide clear transparency regarding the utilization of creators' works for training purposes.³⁰⁸ Although it may serve as a temporary measure, the ultimate goal is to establish universally accepted standards.³⁰⁹ In this sense, quick action seen vital to prevent prolonged uncertainty and maintain the legitimacy of the EU's legal framework for using copyrighted works in ML training.³¹⁰

³⁰⁷ Keller and Warsø, "Defining Best Practices for Opting out of ML Training," 14.

³⁰⁸ "Policy Paper #15 on Using Copyrighted Works for Teaching the Machine," 4.

³⁰⁹ Keller, "Generative AI and Copyright."

³¹⁰ Keller.

Chapter Five: The Implementation of the Articles

In the ongoing implementation of the directive, 11 Member States, failing to notify the European Commission of the necessary transposition measures under CDSM Directive, faced referral to the CJEU.³¹¹ Overall regarding the implementation process in the member states, it can be said that even though TDM exceptions caused debates in discussions, they haven't been a big focus.³¹² It is suggested that this might be because of the language in Articles 3 and 4 of the Directive is clear, providing an easy template for countries to follow.³¹³ Therefore the majority of member states, not having rules for TDM before, have simply added these new exceptions to their copyright laws. In this sense, France and Germany are worth examining since both of them presents different backgrounds in terms of transposition of TDM exceptions.

5.1- Germany

In order to transpose CDSM to national law, Germany recently passed a major copyright law reform on June 7, 2021, titled “*zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes*” (adapting copyright law to the requirements of the digital single market).³¹⁴ The legislation includes provisions that exempt TDM from copyright restrictions, which is a crucial technology for ML and AI (Sections 44b, 60d UrhG).³¹⁵ In the light of this legislation, Section 44b UrhG, serving as the foundational provision, now opens TDM to a broader audience, aligning with Article 4 of the CDSM.³¹⁶ According to 44b, the rightsholders must not reserve

³¹¹ “11 Member States Referred to the Court of Justice of the EU.”

³¹² Paul Keller, “DSM Implementation: What Is Happening with Regards to the Other Provisions,,” February 15, 2021, <https://eu-copyright-implementation.info/en/comments-on-implementation/dsm-implementation-what-is-happening-with-regards-to-the-other-provisions/>.

³¹³ Keller.

³¹⁴ “Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes,” Bundesministerium der Justiz, accessed January 11, 2024, https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2020_Gesetz_Anpassung-Urheberrecht-dig-Binnenmarkt.html.

³¹⁵ “Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes,” See: “Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG),” accessed January 11, 2024, https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0328.

³¹⁶ Jan Bernd Nordemann, Jonathan Pukas, and Julian Waiblinger, “The EU DSM Copyright Directive: Implementation in Germany 2021 – Part I,” Kluwer Copyright Blog, January 16,

non-scientific TDM rights, the work must be lawfully accessible, copies for analytic processes must be deleted when no longer needed, and there is no obligation to pay remuneration for TDM use.³¹⁷ In this perspective, commercial TDM follows basic rules in Section 44b.³¹⁸ Meanwhile Section 60d of the legislation extends provisions for scientific research, but research organizations risk losing exemption when collaborating with certain private enterprises.³¹⁹ Unlike Section 44b, Section 60d permits storage for the duration needed for research purposes or quality monitoring of scientific findings.³²⁰

The newly introduced exception and limitation of TDM in German law, mirroring the CDSM Directive, offers an intriguing framework, particularly in light of the “*Freier Werkgenuss*” doctrine in German case law. This doctrine, found out to be very similar to Japanese “*non-enjoyment*” perspective, means simply enjoying a creative work.³²¹ This idea revolves around the joy and satisfaction we gain from intellectual or artistic creations, which is the main reason why copyrights exist. It is argued that TDM shouldn’t be strictly bound by copyright rules because the main purpose of copyright is to support intellectual satisfaction.³²² According to this; if an action doesn’t really contribute to this kind of joy, it might not be that relevant for copyright. Hence, Japanese approach to TDM considered closer to European models rather than US fair use approach.³²³

5.2- France

Despite being the first Member State to partially implement the CDSM, France’s implementation process faced divisions among different parties, resulting in

2023, <https://copyrightblog.kluweriplaw.com/2023/01/16/the-eu-dsm-copyright-directive-implementation-in-germany-2021-part-i/>.

³¹⁷ “Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG).”

³¹⁸ Nordemann, Pukas, and Waiblinger, “The EU DSM Copyright Directive.”

³¹⁹ “Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG).”

³²⁰ “Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG).”

³²¹ Ueno, “The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes – Recent Amendment in Japan and Its Implication,” 152.

³²² Kollár, “Mind If I Mine? A Study on the Justification and Sufficiency of Text and Data Mining Exceptions in the European Union,” 25.

³²³ Senftleben et al., “Ensuring the Visibility and Accessibility of European Creative Content on the World Market,” 72.

an incomplete transposition that missed the EU’s envisioned deadline.³²⁴ The corresponding articles 3 and 4 of CDSM transposed into French Intellectual Property Code as of Article L122-5-3.³²⁵ The TDM exception for scientific research already existed in French law, but it differed from the CDSM by specifying beneficiaries instead of using an umbrella term like “cultural heritage institutions.”³²⁶ Also, as mentioned in Chapter 4, Article 3 of CDSM doesn’t refer to the possibility of collaboration between beneficiaries of the scientific research exception but in Recital 12, it is stated that nonprofit basis would be the condition for possible collaborations. In this context, the French Intellectual Property Code clearly incorporates this condition into the transposed article.³²⁷

On the other hand, the exception for commercial purposes is a new addition following the CDSM.³²⁸ It aligns with the Directive by allowing copyright holders to opt out of this exception without providing a justification, and they can do so by simply noting it in the file’s metadata.³²⁹

In response to the new exception for commercial purposes, various organizations have initiated the implementation of the opt-out clause. Notably, the French Publishers’ Association (Syndicat National de l’Édition) introduced this exception to its members emphasizing the significance of data in the AI learning process, that the exception allows AI operators to extract content, even if it includes

³²⁴ “CDSM Implementation Resource Page – CREATe,” accessed January 12, 2024, <https://www.create.ac.uk/cdsm-implementation-resource-page/#consultations-transpositions>.

³²⁵ “Article L122-5-3 - Code de La Propriété Intellectuelle - Légifrance,” accessed January 12, 2024, https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044363192#.

³²⁶ “Ouvrir La Science - La Fouille de Textes et de Données à Des Fins de Recherche : Une Pratique Confirmée et Désormais Opérationnelle En Droit Français,” accessed January 13, 2024, <https://www.ouvrirlascience.fr/la-fouille-de-textes-et-de-donnees-a-des-fins-de-recherche-une-pratique-confirnee-et-desormais-operationnelle-en-droit-francais/>.

³²⁷ “Un Décret Précise Les Exceptions En Vue de La Fouille de Textes et de Données,” accessed January 13, 2024, <https://actualitte.com/article/106646/droit-justice/un-decret-precise-les-exceptions-en-vue-de-la-fouille-de-textes-et-de-donnees>.

³²⁸ “🌐 DSM Directive Implementation Tracker,” accessed January 13, 2024, <https://www.notion.so/DSM-Directive-Implementation-Tracker-361cfac48e814440b353b32692bba879>.

³²⁹ “Article L122-5-3 - Code de La Propriété Intellectuelle - Légifrance.”

copyrighted material accessible on the internet.³³⁰ In this sense, French Publishers' Association introduced rightsholders right to retain the option to oppose this through an "opt-out" mechanism and to facilitate this, makes opt-out tool available for its members and recommends to utilize and exercise this option.³³¹

Furthermore, the Society of Authors, Composers and Publishers of Music, (SACEM, Société des Auteurs, Compositeurs et Éditeurs de Musique), which has 427.500 clients who can use 96 millions of works, decided to exercise the right to 'opt-out' under the new legislation regarding TDM.³³² In this regard, entities seeking to utilize SACEM's works for AI training must secure prior authorization, reflecting SACEM's commitment to negotiated and transparent conditions. Contrary to the misconception that this decision may hinder AI development, Cécile Rap-Veber, CEO of Sacem, emphasizes that SACEM actively supports the advancement of AI, and exercising its' opt-out right to carefully strike a balance between the rights of the creators and publishers it represents and the objectives of the AI field.³³³ Apparently, opt out trend escalates within the media sector as well, since entities like Radio France and TF1 (TV channel) are prohibiting OpenAI's GPTBot from accessing their data, indicating a broader shift towards firmer controls on utilizing data for AI training without explicit consent.³³⁴ One of the Directors of a news website explained the reason behind this exercise of right was preventing other businesses, that trains AI from deriving value from their content without reciprocity.³³⁵ Considering all these changes together, situation indicates highlights a growing awareness of the importance of protecting copyrighted content in the digital environment and emphasizes the need

³³⁰ "Une Clause-Type Pour s'opposer à La Fouille de Textes et de Données Par Les Intelligences Artificielles - Syndicat National de l'édition," accessed January 13, 2024, <https://www.sne.fr/actu/une-clause-type-pour-sopposer-a-la-fouille-de-textes-et-de-donnees-par-les-intelligences-artificielles/>.

³³¹ "Une Clause-Type Pour s'opposer à La Fouille de Textes et de Données Par Les Intelligences Artificielles - Syndicat National de l'édition."

³³² "Sacem, in Favour of Virtuous, Transparent, and Fair AI, Exercises Its Right to Opt-out - SACEM," accessed January 13, 2024, <https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out>.

³³³ "Sacem, in Favour of Virtuous, Transparent, and Fair AI, Exercises Its Right to Opt-out - SACEM."

³³⁴ Julia Tar, "Several French Media Block OpenAI's GPTBot over Data Collection Concerns," [www.euractiv.com](https://www.euractiv.com/section/artificial-intelligence/news/several-french-media-block-openais-gptbot-over-data-collection-concerns/), August 29, 2023, <https://www.euractiv.com/section/artificial-intelligence/news/several-french-media-block-openais-gptbot-over-data-collection-concerns/>.

³³⁵ Tar.

for regulations that balance creators' interests with the possibilities offered by advanced technologies.³³⁶

³³⁶ Aude Malharbe, "L'évolution du droit d'auteur face aux défis des intelligences artificielles génératives," Conseil National Des Barreaux, accessed January 12, 2024, <https://consultation.avocat.fr/blog/aude-malherbe/article-170740-l-evolution-du-droit-d-auteur-face-aux-defis-des-intelligences-artificielles-generatives.html>.

Chapter Six: The regime of the TDM outside of the EU - United States and Japan

In numerous jurisdictions outside the EU, TDM activities doesn't necessitate authorizations, owing to legal or judicial developments that promote technological innovation.³³⁷ Because regulating legal aspects of technology is difficult. The difficulty arises from the natural mismatch between the speed of technological advancements and the comparatively slow procedures of legislating and putting laws into practice.³³⁸ In this sense, despite careful drafting and heated debates, legal provisions can quickly become outdated during the time it takes for laws to be enacted and implemented.

Several legal systems beyond the EU, seen as competitors in technology, creativity, and culture, have adopted favorable approaches for technological development. These approaches include open and flexible standards, judicial construction of users' rights, or the implementation of a dedicated TDM limitation for any purpose. Thus, the implication is that the legislative choices made in Articles 3 and 4 may not be as favorable to technological development as those observed in comparable legal systems outside the EU.³³⁹

In this sense, the United States and Japan are worth examining because, despite belonging to different copyright traditions, they both showcase effort to robust creative industries and competitive high-tech sectors in the field of AI.³⁴⁰

6.1- Legal Framework in the United States

In the realm of copyright law, especially regarding the TDM and ML activities, the United States is known for being more flexible with the doctrine of "fair use" in

³³⁷ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 687.

³³⁸ Gerrish and Skavlan, "European Copyright Law and the Text and Data Mining Exceptions and Limitations," 66.

³³⁹ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 687.

³⁴⁰ Senfleben et al., "Ensuring the Visibility and Accessibility of European Creative Content on the World Market," 72.

allowing TDM activities with copyrighted materials compared to the EU.³⁴¹ This makes the United States an appealing choice for companies wanting to develop AI tools.³⁴² This chapter explores the regulatory environment that makes the United States an attractive hub for TDM innovation.

Fair use is a legal principle that permits specific uses of copyrighted materials without requiring permission or compensation to the copyright owner.³⁴³ It was originally developed by the courts and later codified into the Copyright Act of 1976.³⁴⁴ Section 107 of this act specifies that copying a copyrighted work for purposes like scholarship or research is not considered copyright infringement.³⁴⁵ However, it's important to note that this provision does not make such uses automatically fair. Four essential factors are taken into account to determine fair use: “*the purpose and nature of the use, the nature of the copyrighted work, the amount of the work used, and the impact of the use on the potential market for the copyrighted work.*”³⁴⁶ In essence, fair use seeks to strike a balance between the rights of the copyright owner and the rights of users to utilize copyrighted material for specific purposes, such as research or scholarship.

A rather recent and famous take on the fair use doctrine was at the legal dispute involved several universities and officials collaborating with Google on the Google Books project, leading to the creation of HathiTrust Digital Library. In this case known as Authors Guild, Inc. v. HathiTrust (755 F.3d 87 (2d Cir. 2014)) with a vast collection of over ten million works spanning centuries and languages, HathiTrust Digital Library served various purposes, including full-text searching for the public, access for patrons with print disabilities, and replacement of lost library copies.³⁴⁷ The plaintiffs,

³⁴¹ Senfleben et al., 72.

³⁴² Quintais, “Generative AI, Copyright and the AI Act.”

³⁴³ “U.S. Copyright Office Fair Use Index,” accessed December 16, 2023, <https://www.copyright.gov/fair-use/index.html>.

³⁴⁴ Victor S. Dorokhin, “The Fair Use Doctrine in U.S. Copyright Law,” *Medium* (blog), October 16, 2023, <https://medium.com/@victordorokhin1/the-fair-use-doctrine-in-u-s-copyright-law-e65c98049946>.

³⁴⁵ “17 U.S. Code § 107 - Limitations on Exclusive Rights: Fair Use,” LII / Legal Information Institute, accessed December 16, 2023, <https://www.law.cornell.edu/uscode/text/17/107>.

³⁴⁶ “U.S. Copyright Office Fair Use Index.”

³⁴⁷ “Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014),” *U.S. Copyright Office Fair Use Index*, n.d., <http://copyright.gov/fair-use/index.html>.

comprising individual authors and authors' associations, contested the fair use defense granted to the defendants by the district court.³⁴⁸ In conclusion, the court's decision affirmed the fair use of the HathiTrust Digital Library's creation of a full-text searchable database, deeming it transformative and essential to public services.³⁴⁹ The court emphasized the lack of harm to traditional markets and the reasonable necessity of retaining text and image copies for accessibility.³⁵⁰

Afterwards this case and many alike,³⁵¹ it has been said that TDM researchers in the United States have secured a hard-won fair use right, allowing them to maintain reproducibility copies of crucial research articles and data.³⁵² Though it should be kept in mind that the fact that fair use is a common argument in certain conditions, it doesn't mean that it cannot be questioned. In this sense, even though the fair use argument is still strong, the "free riding" nature of ML systems are heavily criticized.³⁵³ Regardless, the recognition of fair use reflects the dedication of researchers and the adaptability of US copyright law to meet the evolving needs of the scientific community.

Comparatively, the competitive advantage of fair use in the United States becomes apparent when contrasted with the EU's CDSM Directive. Fair use's flexibility and focus on public benefits for various uses, especially those driving innovation, stand in contrast to the CDSM Directive's restrictions on TDM exceptions.³⁵⁴ The Directive's reliance on "lawful access" and conditional exceptions highlights the need for a balanced and flexible approach to copyright exceptions that encourages innovation while serving the public interest.

³⁴⁸ "Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014)."

³⁴⁹ "Authors Guild, Inc. v. HathiTrust, No. 12-4547 (2d Cir. 2014)," Justia Law, 18, accessed December 16, 2023, <https://law.justia.com/cases/federal/appellate-courts/ca2/12-4547/12-4547-2014-06-10.html>.

³⁵⁰ "Authors Guild, Inc. v. HathiTrust, No. 12-4547 (2d Cir. 2014)," 22.

³⁵¹ Menachem Kaplan, Marissa Yu, and Luc Von Danwitz, "The Fair Use Defense for Generative AI Tools After Warhol V. Goldsmith," Passle, June 16, 2023, <https://blog.freshfields.us//post/102ih2k/the-fair-use-defense-for-generative-ai-tools-after-warhol-v-goldsmith>; Pamela Samuelson, "Regulating Technology Through Copyright Law: A Comparative Perspective," *European Intellectual Property Review*, no. 42 (2020): 6, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635094.

³⁵² Carroll, "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful," 936.

³⁵³ Lemley and Casey, "Fair Learning," 139–52.

³⁵⁴ Carroll, "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful," 958.

Overall, this approach in copyright law distinguishes itself as a strong supporter of scientific progress, creating a science-friendly environment that positions the United States as a competitive force in TDM research.³⁵⁵ This advantage highlights the crucial need to balance copyright protection with the imperative of promoting scientific progress.

6.2- Legal Framework in Japan

Japanese Copyright Law, introduces a new perspective to the TDM activities: “non-enjoyment”.³⁵⁶ With this perspective, the question asked is simple: Why allow TDM without the intention of “enjoyment”? Typically, copyright law protects works that bring intellectual or emotional enjoyment, like music or literature.³⁵⁷ However, TDM, which is a computer program that doesn’t seek enjoyment, doesn’t harm the interests copyright law aims to protect. Japan’s decision to permit TDM for non-enjoyment purposes reflects a nuanced understanding of copyright principles, balancing technological progress with copyright holder rights.

In 2016, Japan embraced AI as a key technological pillar to establish a super-smart society, also known as “Society 5.0.”³⁵⁸ This vision, emphasizing the transformative impact of technological progress on societal dynamics, categorizes human societies into distinct stages: Society 1.0, focused on hunting and gathering; Society 2.0, centered around agriculture and nation-building; Society 3.0, driven by industrialization; and Society 4.0, characterized as an information society.³⁵⁹ As part of this evolution, the CEO of a prominent AI development company in Japan highlights the importance of AI, stating that in the future, it will be considered essential infrastructure, alongside electricity, gas, and water.³⁶⁰ Society 5.0, building upon the

³⁵⁵ Carroll, 936.

³⁵⁶ “Copyright Law of Japan | Copyright Law of Japan | Copyright Research and Information Center CRIC,” accessed December 15, 2023, <https://www.cric.or.jp/english/clj/cl2.html>.

³⁵⁷ Dermawan, “Text and Data Mining Exceptions in the Development of Generative AI Models,” 12.

³⁵⁸ “Artificial Intelligence / The Government of Japan - JapanGov -,” accessed December 15, 2023, https://www.japan.go.jp/tomodachi/2018/spring2018/artificial_intelligence.html.

³⁵⁹ Mayumi Fukuyama, “Society 5.0: Aiming for a New Human-Centered Society,” *Japan Spotlight*, August 2018, 47–50.

³⁶⁰ “Artificial Intelligence / The Government of Japan - JapanGov -.”

advancements of Society 4.0, envisions a future where AI integration is pivotal for driving economic growth and effectively addressing societal challenges.

Japan's unwavering commitment to technological progress and innovation is evident in its' "non-enjoyment" clauses; Article 30-4 standing out as a crucial provision for TDM. Article 30-4 is characterized as the "*broadest TDM exception in the world*," allowing use works for TDM objectives, including extraction, comparison, classification, or other statistical analyses.³⁶¹ Notably, this exception covers both commercial and noncommercial purposes, is not subject to rightsholders' reservations, permits exploitation by any means, and does not require lawful access. Japan's progressive approach has effectively positioned the country as a new "*Paradise for AI and machine learning*" providing legal certainty and flexibility to AI innovators and establishing itself as a global leader in fostering a conducive environment for technological advancements.³⁶²

As mentioned in Chapter 1, international copyright treaties require countries to follow a three-step test for copyright exceptions in their national laws.³⁶³ In this perspective, the Japanese Copyright Act's Article 30-4 aligns with these standards by avoiding conflicts with normal work use and not unfairly impacting copyright holders.³⁶⁴ An interesting consequence of this is the absence of a right of payment for authors in commercial TDM activities. Normally, one might expect payment for their work being used commercially, but the Japanese Government argues that exclusive and payment rights aren't necessary for TDM, as these activities, even when commercial, don't harm the chances of copyright holders getting compensated.³⁶⁵ This approach reflects a shift in copyright dynamics, prioritizing innovation without compromising creators' interests.

³⁶¹ Ueno, "The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication." 149.

³⁶² Ueno, 149.

³⁶³ Article 9(2) World Intellectual Property Organization, *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)*; Article 13 "WTO | Intellectual Property (TRIPS) - Agreement Text - Standards," accessed June 20, 2024, https://www.wto.org/english/docs_e/legal_e/27-trips_04_e.htm#.

³⁶⁴ Ueno, "The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication," 150.

³⁶⁵ Ueno, 150.

Chapter Seven: Legal Framework in Türkiye

7.1- Current Situation

As of today, there is no provision in Turkish law corresponding to the exceptions for TDM, and there is no draft law being discussed on this subject.³⁶⁶ Türkiye, being an EU candidate for 25 years, receives yearly reports on chapters for harmonization with the EU acquis.³⁶⁷ In this sense, 2023 report of the EU Commission states that the current copyright law in Türkiye has not undergone complete modernization to offer adequate protection in accordance with technological advancements, and Türkiye is not yet in compliance with the CDSM Directive.³⁶⁸ Within this background, the TDM activities carried out would possibly violate the reproduction and transmission rights of authors and database owners' rights in accordance with the current legislation, Law on Intellectual and Artistic Works, Law No. 5846 dated 5.12.1951 (LIA).³⁶⁹ With the rapid advancements in AI, ML, GenAI, inevitably in connection with TDM, and the growing global regulatory landscape, Türkiye is expected to join the ranks of states that regulate TDM one way or another. In this respect, the doctrine acknowledges the need for a legal framework to balance

³⁶⁶ "Türkiye Büyük Millet Meclisi," accessed January 16, 2024, <https://www.tbmm.gov.tr/Yasama/Kanun-Teklifleri-Sonuc>; "Mevzuat Bilgi Sistemi," accessed January 16, 2024, <https://mevzuat.tbmm.gov.tr/>. (as a result of queries made in search engines of legislation and legislative proposals)

³⁶⁷ See for the long period of candidate status: "The Fiction That Turkey Is a Candidate to Join the EU Is Unravelling," *The Economist*, accessed January 16, 2024, https://www.economist.com/europe/2021/08/26/the-fiction-that-turkey-is-a-candidate-to-join-the-eu-is-unravelling?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18151738051&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAiA75itBhA6EiwAkho9e-X2cjEaW-PBzfCdEIXWVFHQE6ywQNJ-ccOVGWgPNUWX3rXaKAVb_RoCJTAAvD_BwE&gclsrc=aw.ds.

³⁶⁸ "Commission Staff Working Document Türkiye 2023 Report Accompanying the Document Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 2023 Communication on EU Enlargement Policy" (2023), 88, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=SWD:2023:696:FIN>.

³⁶⁹ in english: "WIPO Lex- Law No. 5846 of December 5, 1951, on Intellectual and Artistic Works (as Amended up to Law No. 6552 of September 10, 2014)," WIPO, accessed January 16, 2024, <https://www.wipo.int/wipolex/en/text/440051>; "Fikir ve Sanat Eserleri Kanunu," Pub. L. No. Yayınlandığı R.Gazete : Tarih : 13/12/1951 Sayı : 7981 Yayınlandığı Düstur : Tertip : 3 Cilt: 33 Sayfa : 49, 5846 (1951), <https://www.mevzuat.gov.tr/mevzuatmetin/1.3.5846.pdf>.

the interests of those involved in these activities, the broader public interest, and the rights of authors whose works are utilized in these processes.³⁷⁰

7.2- Legislative Background

According to the current legislative background, copyright holders have financial and moral rights. The financial rights of the copyright holder include the rights to process, reproduce, distribute, represent, and communicate the work to the public using sound, video, or image transmission tools.(LIA Art.20) Meanwhile, the moral rights of the copyright holder are defined as the right to present the work to the public, the right to determine the name of the work, the right to prohibit modifications to the work, and the rights that the copyright holder can exercise against owners and possessors.(LIA Art.14) In this sense, a TDM activity- which is any AI program utilizes TDM-, that consists of copying, extracting, and recombining data, would violate Article 21 and 22 of the LIA.

In relation with databases, Article 6.11 and Additional Article 8 of the Law would be in question of violation. Databases receive dual protection under the law.³⁷¹ Firstly, they are protected by copyright, as outlined in Article 6.11 of LIA. Secondly, database producers are granted sui generis rights as compensation for their investment in creating the database, according to additional Article 8 of the same law. To be considered as a work, according to Article 6.11, a database must select and compile data, turning it into usable information.³⁷² The database can contain different types of

³⁷⁰ Osman Gazi Güçlütürk, “Türk Hukukunda Makine Öğrenmesine Dayalı Yapay Zekada Verinin Hukuka Uygun Şekilde Kullanılması” (Phd Thesis, İstanbul, Galatasaray Üniversitesi, 2021), 230, <https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=DUkPWcgZfXOGAXLTxfIw&no=dEVj8z85tRmLg-yXXWxLtA>; Mustafa Halil Çonkar, “Dijital Tek Pazarda Telif Hakkı ve Bağlantılı Haklar,” in *Türk hukukunun Avrupa Birliği hukukuna uyumu özel hukuk: acquis communautaire’in alınması: açıklamalar, değerlendirmeler, öneriler = Approximation of Turkish law with European Union law private law: transposition of acquis communautaire: remarks, assessments, suggestions*, ed. Arslan Kaya et al., İstanbul Üniversitesi yayınları, no 5269 (İstanbul: İstanbul Üniversitesi Hukuk Fakültesi, 2020), 704; Saniye Ezgi Ercan, “Avrupa Birliği Dijital Tek Pazarda Telif Hakkı Yönergesi ile Öngörülen İstisnalar Çerçevesinde Türk Fikir ve Sanat Eserleri Kanununun Değerlendirilmesi” (Masters Thesis, Ankara, İhsan Doğramacı Bilkent Üniversitesi, 2022), 97, <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>.

³⁷¹ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2.Baskı, vol. I (Ankara: Seçkin Yayıncılık, 2014), 173–74.

³⁷² Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:173–74.

data and materials, and how it's compiled doesn't matter.³⁷³ Also, it needs to cover a big part of the data available in its field.³⁷⁴ Databases should be flexible, so changes can be made without losing coherence.³⁷⁵

Sui generis databases, lacking creative characteristics, aren't covered by typical copyright laws.³⁷⁶ Additional Article 8 grants sui generis right to database producers, separate from copyright, recognizing their substantial investment in creating and maintaining databases.³⁷⁷ The law defines the database producer as someone who invests substantially in the creation, verification, presentation, and substantial proportion of the content of the database in terms of quality and quantity. The importance of this protection can be seen if such an investor is not protected, the creation of investment-dependent databases cannot be realized.³⁷⁸ In an example given by Tekinalp, it is suggested that creating a database related to historical artifacts, vegetation, clean water sources, or minerals in Türkiye requires billions of lira in investment.³⁷⁹ Individuals investing in these areas also want to see results from their investments, hence the reason for granting rights to the database producer is to ensure this security. This protection lasts for 15 years and grant producers control over certain database actions such as content transfer and reuse restrictions.³⁸⁰ Significant changes to the database trigger a new protection period, and even partial use of the content may constitute infringement.³⁸¹

³⁷³ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2.Baskı, vol. I (Ankara: Seçkin Yayıncılık, 2014), 173–74.

³⁷⁴ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:173–74.

³⁷⁵ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:173–74.

³⁷⁶ İbrahim Hulusi Külahcı, “Elektronik Veri Tabanlarının Fikir ve Sanat Hukuku Kapsamında Korunması” (Masters Thesis, İstanbul, İstanbul Sabahattin Zaim Üniversitesi, 2019), 138, <https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=vjszP7PzV0HebcjFEvDfwICZTYGOQdmTgR9qC193FdODp4hXqNt2IJRfGxIMUjce>; Yavuz Selim Şener, “Fikri Mülkiyet Hukukunda Dijital Veri Tabanlarının Korunması” (Masters Thesis, İstanbul, İstanbul Kültür Üniversitesi, 2013), 89, <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>.

³⁷⁷ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2.baskı, vol. II (Ankara: Seçkin Yayıncılık, 2014), 3008–9.

³⁷⁸ Ünal Tekinalp, *Fikri Mülkiyet Hukuku*, 5. Basım (İstanbul: Vedat Kitapçılık, 2012), 141.

³⁷⁹ Ünal Tekinalp, 141.

³⁸⁰ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, II:3008–9; Şener, “Fikri Mülkiyet Hukukunda Dijital Veri Tabanlarının Korunması,” 89.

³⁸¹ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, II:3008–9.

In this chapter, for a thorough examination of the legal framework, initial focus will be directed towards analyzing Articles 21 and 22, followed by an in-depth exploration of the provisions concerning to databases. Subsequently, an investigation into the legal remedies available to rights holders in instances of infringement will be conducted.

As mentioned in Chapter One, in a TDM activity, three steps occur: data accessibility, content extraction/copying, and pattern identification. First step, data accessibility, will be further explored below in relation to legal methods for accessing databases and their associated rights. Other than within a database, accessing data is possible through varying online sources as well. In this scenario, use of copyrighted works should be evaluated under second step. The third step, pattern identification, represents the transformative essence of TDM, highlighting its capacity to extract valuable insights from data while minimizing dependence on specific content.³⁸² Consequently, it is often argued that outcome of the TDM does not essentially infringe upon IP rights, as factual information typically falls outside the scope of IP protection.³⁸³ So, the second step, extracting and copying content, even if temporary, could breach the rightsholders' right to reproduce their work in any manner, including copying, recording, and computer program activities, as specified in Article 21, "Right to Process" and Article 22, known as the "Right of Reproduction".

According to Article 21 of the LIA, authors have the exclusive right to process their works, although processing for personal use is generally allowed.³⁸⁴ Third parties may process the work with the author's permission, but this must be within the scope of that permission.³⁸⁵ The transfer of all rights over a work does not automatically

³⁸² Carroll, "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful," 895.

³⁸³ TRIPS art. 9(2): "Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such." "WTO | Intellectual Property (TRIPS) - Agreement Text - Standards"; WCT art. 2: "Copyright protection extends to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such." "WIPO Copyright Treaty"; Rosati, *Handbook on the Economics of Copyright*, 51; Geiger, Frosio, and Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," 4.

³⁸⁴ Cahit Suluk, Rauf Karasu, and Temel Nal, *Fikri Mülkiyet Hukuku*, 7th ed. (Ankara: Seçkin Yayıncılık, 2023), 116; Süleyman Emre Özdemir, "Eseri Çoğaltma Hakkının Cezai Himayesi ve Şahsen Kullanma Meselesi," *Marmara Üniversitesi Hukuk Fakültesi Hukuk Araştırmaları Dergisi* 28, no. 1 (June 30, 2022): 170–83, <https://doi.org/10.33433/maruhad.1094016>.

³⁸⁵ Suluk, Karasu, and Nal, *Fikri Mülkiyet Hukuku*, 116.

transfer the processing right unless explicitly stated otherwise.³⁸⁶ Adaptation of computer programs falls under the processing right according to the Art. 38 which regulates the personal use exception, but TDM applications are not possibly covered by personal use exceptions.

Transforming an existing work into a new form through creative effort results in a valuable product that enriches culture.³⁸⁷ The work resulting from the processing of a work should resemble the original to some extent but can still be recognized as distinct, even if it retains connections to the original.³⁸⁸ For instance, translating a work into a different form, like turning a novel into a film, involves maintaining the essence of the original while adding new elements.³⁸⁹ Creativity is crucial for adaptation rights, ensuring the contribution reflects the adapter's characteristics, as mandated by law. (LIA Art.1/B, c)³⁹⁰

The result of processing a work can be considered a new work under copyright law, even without permission from the original creator, as long as it reflects the original's characteristics.³⁹¹ However, unauthorized use of the processed work outside of personal circles violates the original creator's rights, requiring permission from them for public presentation or economic benefit.³⁹² This is outlined in the LIA, allowing processing owners to use their rights only with the original creator's consent. (Art 20.2)³⁹³ On the contrary, without making it public, it is possible to process a previously published work without permission.³⁹⁴ In this scenario there is no need to protect the author against such processing actions because neither their financial nor moral interests are harmed.³⁹⁵

³⁸⁶ Article 55 "WIPO Lex- Law No. 5846 of December 5, 1951, on Intellectual and Artistic Works (as Amended up to Law No. 6552 of September 10, 2014)."

³⁸⁷ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:737.

³⁸⁸ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:737.

³⁸⁹ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:739.

³⁹⁰ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:739.

³⁹¹ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:745.

³⁹² Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:748.

³⁹³ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:748.

³⁹⁴ Ünal Tekinalp, *Fikri Mülkiyet Hukuku*, 183.

³⁹⁵ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:745.

Reproduction, as stated in Article 22, covers a broad range of activities, including the generation of temporary copies, which are the second step of the TDM activities as mentioned above. Whether through recording, duplication, or any other means, reproduction aims to make the work tangible for later use. Therefore, unauthorized reproduction infringes upon the creator's rights, except for specific personal use exceptions outlined in the law.³⁹⁶ Reproduction rights serve as the cornerstone of copyright law, providing the foundation from which all other rights stem.³⁹⁷ While the law does not offer a precise definition of reproduction, it interprets the concept broadly, which allows various forms of reproduction to safeguard the rights of creators.³⁹⁸

In reproduction, the used method has no importance—it could be handwritten, printed, digital, or analog.³⁹⁹ The important thing is capturing the essence of the original. Yet, fine art's worth comes from its uniqueness, so copying it without using the original tools wouldn't be quite reproduction; it might be seen as a new piece or an adaptation instead.⁴⁰⁰ Reproduction involves exact copying, therefore unauthorized imitation would be piracy and infringement. Additionally, reproduction rights extend to actions related to computer programs, such as loading them onto a computer, displaying them, running them, and storing them temporarily.⁴⁰¹ They also extend to various digital activities, such as scanning documents, uploading files, downloading content, and sharing files over digital platforms.⁴⁰²

³⁹⁶ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:817.

³⁹⁷ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, I:817.

³⁹⁸ Ünal Tekinalp, *Fikri Mülkiyet Hukuku*, 184–86; Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:817; Suluk, Karasu, and Nal, *Fikri Mülkiyet Hukuku*, 116.

³⁹⁹ Suluk, Karasu, and Nal, *Fikri Mülkiyet Hukuku*, 116.

⁴⁰⁰ Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:823.

⁴⁰¹ Ünal Tekinalp, *Fikri Mülkiyet Hukuku*, 185; Levent Yavuz, Türkay Alıca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:841–43; Suluk, Karasu, and Nal, *Fikri Mülkiyet Hukuku*, 116; Abdullah Çelik, “Fikir ve Sanat Eserleri Üzerindeki Çoğaltma ve Yayma Haklarının İhlali, İhlalin Sonuçları” (Phd Thesis, İzmir, Dokuz Eylül Üniversitesi, 2010), 81–97, <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>.

⁴⁰² Çelik, “Fikir ve Sanat Eserleri Üzerindeki Çoğaltma ve Yayma Haklarının İhlali, İhlalin Sonuçları,” 81–97.

It has been stated at the Article 22, that the right to reproduce the work in whole or in part belongs to the owner of the work. And regardless of whether its permanent or for limited time, reproduction is reproduction.⁴⁰³ For example, placing the work on the internet is such a reproduction. Because Article 22.1 of the LIA also states that the right to make temporary reproductions of the work belongs to the owner of the work. Moreover, the act of reproducing a work without making it public still violates the owner's reproduction right, regardless of whether the work has been shared with the public.⁴⁰⁴ Because this right is separate from the right to distribute the work publicly. Even if unauthorized copies are not sold, the owner can demand their destruction or receive compensation not exceeding the production cost. (LIA Art. 68)

In this sense, its suggested that temporary or unintentional reproductions made during the transmission or legal use of a work through technology, which don't have any economic importance, shouldn't considered a violation of the copyright holder's reproduction rights, even if not explicitly stated in the law.⁴⁰⁵ It can be discussed that whether or not this evaluation reflects the TDM process or not but it's clear that the outcome of the TDM possibly have significant economic importance, hence it would violate reproduction rights in any case.

First step of the TDM, access to the data concerns the database rights that regulated in Article 6.11 and Additional Article 8 of the LIA. As mentioned above, there is two distinct protection systems have been adopted for databases. The concept of database is defined in Turkish law with the addition of the 11th clause to Article 6 of LIA in 1995. Databases with original characteristics belonging to their owner are regulated in Article 6.11 of the Law and copyright protection is granted to these original databases. For databases that do not have original characteristics, sui generis right protection has been granted by an additional article (8) to the LIA in 2004 in line with the EU Directive 96/9.⁴⁰⁶

⁴⁰³ Suluk, Karasu, and Nal, *Fikri Mülkiyet Hukuku*, 116.

⁴⁰⁴ Levent Yavuz, Türkay Alca, and Fethi Merdivan, *Fikir ve Sanat Eserleri Kanunu Yorumu*, 2014, I:825.

⁴⁰⁵ Levent Yavuz, Türkay Alca, and Fethi Merdivan, I:846; Pelin Karaaslan, "Fikir ve Sanat Eserlerinin Teknik Gereklik Nedeniyle Geçici Çoğaltılması," *Ticaret ve Fikri Mülkiyet Hukuku Dergisi* C:8, no. S:1 (2022): 36.

⁴⁰⁶ Ünal Tekinalp, *Fikri Mülkiyet Hukuku*, 141.

It's worth noting that the TDM exceptions introduced to the sui generis database right by the CDSM Directive Article 3 do not have a counterpart in Turkish sui generis database right. According to current legislation, the database producer, who invests substantially in content curation, has the right to control the content transfer to another platform, with exceptions outlined in the law. (Add.Art. 8) Referred as the right to prohibit extraction, aims to prevent the transfer of a database's content to another medium without authorization. Electronic transformation of databases enables easy reproduction, transmission, copying, and storage, often without the author's consent. Unauthorized actions in this regard violate the rightsholders extraction or reproduction right. A legal case involving the unauthorized use of a database program illustrates this, with the court ruling against the infringing party for installing the database on other computers without permission.⁴⁰⁷ Decision explains that utilizing a computer program and database without a license, specifically uploading, and storing the program on another computer, constitutes a violation of the right to reproduce under paragraph 3 of Article 22 of the LIA, leading to compensation for the rightful owner of the database. This right resembles the reproduction right granted to authors, safeguarding the efforts and expenses of database producers.⁴⁰⁸ Additionally, they have control over content dissemination, sale, rental, or public communication, except where public safety, administrative, or judicial proceedings require otherwise. (LIA Add.Art.8)

Referred as the right to prohibit re-utilization aims to prohibit the distribution or public transmission of a significant portion of a database's content without permission, akin to the distribution and communication rights in copyright law.⁴⁰⁹ Exceptions for educational and scientific research purposes have been regulated in Article 34 of the LIA. However, the LIA does not address exceptions regarding electronic databases comprehensively, the differentiation in exceptions between

⁴⁰⁷ Yargıtay 11.HD. T.11/05/2001, E.2001/1084, K.2001/4344, as cited in Çelik, "Fikir ve Sanat Eserleri Üzerindeki Çoğaltma ve Yayıma Haklarının İhlali, İhlalin Sonuçları," 86; Külahcı, "Elektronik Veri Tabanlarının Fikir ve Sanat Hukuku Kapsamında Korunması," 63.

⁴⁰⁸ Şener, "Fikri Mülkiyet Hukukunda Dijital Veri Tabanlarının Korunması," 124.

⁴⁰⁹ Şener, 125.

electronic and non- electronic databases doesn't exist in LIA whereas the Article 9 of the Directive 96/9 clarifies a distinction.⁴¹⁰

Considering the TDM activities mostly take place on digital databases, highlighting the legal framework surrounding licensing agreements is also crucial when discussing legal access to them. License agreements facilitate the transfer of usage rights for digital databases, specifying terms like exclusivity and duration.⁴¹¹ Users must respect copyright and database structure, as unauthorized reproduction breaches agreements and constitutes copyright infringement.⁴¹² Therefore, compliance with terms and laws is essential for access to online databases, and unauthorized sharing is prohibited.

Within this legal framework, actions that can be taken towards violations differs whether it's a database has a creative characteristic - an original database- or a sui generis database. In this sense, violations such as exact copying of database content, imitation of database structure, and transferring database information to other platforms using digital and technological methods that might occur during TDM activities would have legal consequences. Original databases are considered literary works in the Bern Convention, resulting the protection granted to these products is a form of copyright protection.⁴¹³ Since original databases benefit from copyright protection, if the moral and financial rights of the owners are violated, they can sue for the infringement to be stopped and seek compensation.(LIA Art. 66-70) This lawsuit can be filed together with or separately from cases concerning determining ownership, preventing infringement, and claiming damages.

On the other hand, sui generis database producers have the right to prevent the transfer or reuse of their database content without permission except for cases required for public safety or required by administrative and judicial procedures. (LIA Add. Art. 8) If unauthorized use occurs, producers can seek legal action for infringement, though compensation for material damages is not explicitly addressed in the law. (LIA

⁴¹⁰ K lahcı, "Elektronik Veri Tabanlarının Fikir ve Sanat Hukuku Kapsamında Korunması," 105.

⁴¹¹ Őener, "Fikri M lkiyet Hukukunda Dijital Veri Tabanlarının Korunması," 116; Őirin Ayd ncık, *Fikri haklara iliŐkin lisans s zleŐmeleri* (İstanbul: Arıkan Basım Yayım, 2006), 60.

⁴¹² Őener, "Fikri M lkiyet Hukukunda Dijital Veri Tabanlarının Korunması," 116.

⁴¹³ Mustafa AteŐ, *Fikri Hukukta Eser* (Ankara: Turhan Kitapevi, 2007), 352.

Art.72)⁴¹⁴ Producers can claim compensation based on general legal provisions if the infringer is at fault and can also seek remedies like cessation of infringement.⁴¹⁵ Additionally, they can pursue legal action under unfair competition laws and request the return of unjust enrichment.⁴¹⁶ While moral rights are not explicitly granted, producers can still claim moral damages for the exploitation of their efforts in creating the database.

Apart from violations concerning databases, in cases of infringement regarding the Articles 21&22, actions specified in the LIA may be pursued by rightsholders, allowing for claims such as determination of the infringement, interdiction of the infringement, removal of the illegal action, compensation. In a more detailed examination of the possible legal actions might be pursued against a subject that conducts TDM, Article 66 allows individuals to sue for damages if their moral and financial rights are violated, regardless of fault. In this case factors like the extent of infringement are considered for compensation. Article 67 allows lawsuits for moral rights infringement, specifying conditions for cessation of infringement. Article 68 covers actions for economic rights infringement, including seeking damages or destruction of unauthorized copies. Article 69 enables prevention of potential infringements, while Article 70 allows compensation for damages. Article 71 outlines penalties for various infringements, while Article 72 addresses circumvention of technological protection measures, punishable by up to two years' imprisonment.

7.3 - Challenges and Considerations in TDM and Legal Framework in Türkiye

Before the implementation of CDSM Directive, in its impact assessment report on the modernization of EU copyright rules, the EU Commission highlighted the challenges faced by TDM research in the EU, primarily due to high licensing costs.⁴¹⁷

⁴¹⁴ İbrahim Emre Bayamhoğlu, “Fikir ve Sanat Eserleri Hukukunda Teknolojik Koruma” (Phd Thesis, Istanbul, Marmara Üniversitesi, 2007), 305; Külahcı, “Elektronik Veri Tabanlarının Fikir ve Sanat Hukuku Kapsamında Korunması,” 157.

⁴¹⁵ Mustafa Ateş, “Veri tabanlarının Hukuki Koruması,” *Ankara Üniversitesi Hukuk Fakültesi Dergisi* 55, no. 1 (2006): 80–81, https://doi.org/10.1501/Hukfak_0000000365; Bayamhoğlu, “Fikir ve Sanat Eserleri Hukukunda Teknolojik Koruma,” 305.

⁴¹⁶ Bayamhoğlu, “Fikir ve Sanat Eserleri Hukukunda Teknolojik Koruma,” 305.

⁴¹⁷ European Commission, Impact Assessment on the modernisation of EU copyright rules.

As detailed in Chapter One, differing provider conditions and licensing practices contribute to the complexity of TDM, resulting in increased expenses for research organizations. One remarkable example of costs show that a leading UK university (University College London) reportedly spends up to GBP 500,000 annually on TDM licenses.⁴¹⁸ While TDM exceptions offer a potential solution, their absence in Türkiye means that licensing costs continue to burden research organizations. The severity of the situation can be understood by the fact that online, free access to the legal precedents of the Supreme Courts wasn't possible. Because they weren't published online by the Court and since the physical search through the written precedents would require long hours of reading hence the need for private databases were born. As a result, making a search through the precedents required licensing of private databases until 2023,⁴¹⁹ even though it is traditionally considered free access in doctrine.⁴²⁰ Therefore, the need for database license agreements remains a significant obstacle to conducting TDM in Türkiye.

The overlapping rights of copyright owners and database creators add complexity, especially in AI applications like creating songs, where the rightsholders numbers multiplies for each component of a song.⁴²¹ This complexity highlights the limitations in the legal framework for TDM activities and becomes an obstacle in the progress of AI development. From technical perspective, copyright holders can implement the machine-readable means such as metadata, robot.txt, website terms and conditions to prohibit TDM activities.⁴²² Considering the awareness on the value of the data is arising, copyright holders might choose not to feed the AI for free without their consent.⁴²³ At this point, Article 4.3 of the CDSM Directive aims to create a

⁴¹⁸ European Commission.

⁴¹⁹ "Yargıtay - Yargıtay İctihat Merkezi Kullanıma Açılmıştır," T.C. Yargıtay Başkanlığı, accessed June 24, 2024, <https://www.yargitay.gov.tr/item/1763/yargitay-ictihat-merkezi-kullanima-acilmistir>.

⁴²⁰ Servet Yetim, "Yargı Kararlarına Erişim Hakkı," *Uyuşmazlık Mahkemesi Dergisi* 4, no. 4 (April 1, 2014), <https://doi.org/10.18771/umdm.90664>.

⁴²¹ Christensen, "A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence," 21.

⁴²² Keller and Warso, "Defining Best Practices for Opting out of ML Training," 7–10.

⁴²³ Kate Knibbs, "Most Top News Sites Block AI Bots. Right-Wing Media Welcomes Them," *Wired*, accessed April 8, 2024, <https://www.wired.com/story/most-news-sites-block-ai-bots-right-wing-media-welcomes-them/>; Benj Edwards, "University of Chicago Researchers Seek to 'Poison' AI Art Generators with Nightshade," *Ars Technica*, October 25, 2023,

balance between protection of the creators and rightsholders and providing a legal base for the developers of ML systems, by offering mechanisms for creators to opt out of their works' use in ML training while ensuring innovation in AI development. Without such regulations, there may be a tendency among copyright holders to restrict AI's ML activities.

Regarding the future legislations that may revolve around TDM exceptions, the presumptions gather around the idea that implementing concerning articles in CDSM Directive while considering the critiques and problems that Member States faced in the transposition process.⁴²⁴ Opting for this choice appears straightforward and familiar, given former efforts to harmonize with the EU.⁴²⁵ Moreover, it aligns logically with the Turkish legal framework in copyright, which bears a closer resemblance to the “droit d'auteur” systems of continental Europe than the Anglo-Saxon systems.⁴²⁶ However, given the significance of potential exceptions that could curtail technological progress, it would be ideal for regulations to align with the practical dynamics of supply and demand in Türkiye, rather than merely adapting regulations elsewhere just as it is. In this context, it is important to address the fact that Türkiye is not in the same stage of development with EU either economically or technologically. It is a crucial factor because these articles criticized heavily on impeding AI technologies, prompting scientists and SMEs to seek more welcoming legal environments for their activities. This means that implementing same rules could have more drastic outcomes in comparing EU. However, choosing not to regulate would also create problems for the right holders, potentially turning Türkiye into a cheap playground for anyone looking to develop an ML system due to the unrestricted access to data. Hence, it is vital for Türkiye to focus on creating a balance between these aspects of a regulation concerning TDM exceptions.

<https://arstechnica.com/information-technology/2023/10/university-of-chicago-researchers-seek-to-poison-ai-art-generators-with-nightshade/>.

⁴²⁴ Güçlütürk, “Türk Hukukunda Makine Öğrenmesine Dayalı Yapay Zekada Verinin Hukuka Uygun Şekilde Kullanılması,” 230; Ercan, “Avrupa Birliği Dijital Tek Pazarda Telif Hakkı Yönergesi ile Öngörülen İstisnalar Çerçevesinde Türk Fikir ve Sanat Eserleri Kanununun Değerlendirilmesi,” 70.

⁴²⁵ For example, this author suggests that same regulations should be adopted in examining Art.15 of the CDSM Directive: Suluk, “Avrupa Birliği Hukukunda Basın Yayıncılara Tanınan Bağlantılı Haklar,” 217–18.

⁴²⁶ Bayamlioğlu, “Fikir ve Sanat Eserleri Hukukunda Teknolojik Koruma,” 97.

It is important to note that this kind of research was beyond the scope of this study. Therefore, the suggestions for future regulation emphasize the importance of conducting a detailed background analysis, as demonstrated by the French example. In order to tailor regulations that address the specific needs of Türkiye, one might consider a study conducted by the Supreme Council for Literary and Artistic Property of the French Ministry of Culture on the transposition of TDM exceptions as an illustrative example.⁴²⁷ It can be seen in the annexes of this report, in order to fully comprehend the challenges and possible suggestions, several questions asked to the contributors which consists of authorities, rightsholders and their representatives, depository organizations, universities and companies in private sector.⁴²⁸ One notable query asked contributors to assess the significance of TDM in their business, quantify related activities with recent examples, and explore the ongoing connection between TDM and AI in the contributor's sector.⁴²⁹ Considering the EU's proactive stance in creating a TDM exception to enhance competitiveness in the evolving technological landscape, as mentioned in Chapter two, and noting that France, as a member state, conducted thorough research during the transposition process into national law, it is reasonable to anticipate that Türkiye should adopt a more thoughtful and suitable approach, rather than settling for the closest option.

⁴²⁷ Report and the Annexes of the report can be found in this link: "Mission du CSPLA sur les exceptions de fouille de textes et de données ('text and data mining')," accessed January 17, 2024, <https://www.culture.gouv.fr/Nous-connaître/Organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA/Mission-du-CSPLA-sur-les-exceptions-de-fouille-de-textes-et-de-données-text-and-data-mining>.

⁴²⁸ "Mission du CSPLA sur les exceptions de fouille de textes et de données ('text and data mining')," 102–5.

⁴²⁹ "Mission du CSPLA sur les exceptions de fouille de textes et de données ('text and data mining')," 67.

Conclusion

This thesis aimed to explore whether recent TDM exceptions in copyright law hinder or foster innovation. To assess this, the crucial elements, regulatory background, and concerning articles were introduced. Emphasizing the intersection of TDM, Big Data, and copyright law in the dynamic realm of intellectual property, the study highlights the potential risks of infringement within big data, as validated by the CJEU in the Infopaq case. This significance amplifies in the context of AI and ML, where TDM navigates extensive data, potentially triggering widespread copyright challenges.

Within this technical background, unresolved copyright-related issues considered as a problem by EU Commission, that they are hindering the progress of TDM in European research, potentially causing a setback.⁴³⁰ This situation raised concerns about EU's global standing as a research destination, affecting the ability of EU universities to attract and retain high-caliber scientists. In order to address these challenges and prevent the EU from falling behind in the field of TDM, mandatory articles of TDM exception implemented with the CDSM. The EU copyright framework, as shaped by the CDSM Directive, described as standing at a "crossroad", facing key decisions influenced by internal contradictions.⁴³¹ These contradictions are rooted in the clash between traditional rules and contemporary digital practices, a global challenge emerging with the advent of digital technologies. The EU's response involved reinforcing copyright protections, extending rights, and restricting free uses, leading to a reduction in such uses and a contraction of the public domain. The second contradiction lies in the EU's limited authority to regulate copyright, relying on internal market mandates.⁴³²

From the assessment of these articles, it can be stated that in this attempt to regulate the single market more horizontally through mandatory articles of Article 3 and 4, potentially falls short in effectively tackling the challenges posed by digital

⁴³⁰ European Commission, Impact Assessment on the modernisation of EU copyright rules.

⁴³¹ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 698.

⁴³² Geiger, Frosio, and Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects," 14.

technologies.⁴³³ For instance, because of this Directive, it can be said that AI sector in EU is at a disadvantaged position against globally, since AI development in EU becomes considerably expensive with requirements for negotiations for licenses covering extensive volumes of works essential as input data.⁴³⁴ With this cost point, AI training becomes more and more problematic. Initially, the higher the price gets, the nature and quality of the data increases. Subsequently, lower budget companies might prefer training their algorithms on free or less expensive data. Which eventually may lead a result that competing dominant AI players will be impossible and inequality prevails the AI sector. Also, transparency of the input data gains more importance since pre trained models might be cheaper therefore might be chosen by developers yet may result obscure, unclear, and inexplicable.⁴³⁵ Therefore it can be said that the rules about TDM in the CDSM Directive might limit AI's full potential. This raises doubts about how well the Directive can handle the connection between copyright and AI in the digital era.

Hence, in the EU, when using works for ML purposes and conducting TDM activities, it is considered wise to either get permission for copyrighted works or choose materials without copyright restrictions, with a focus on open content.⁴³⁶ This is evident in major AI projects that use public domain works, showing a deliberate effort to avoid legal issues.⁴³⁷ As a result, the CDSM Directive appears out of step with the demands of the digital age. Which might be understood as it fails to seize the opportunity to modernize European Copyright Law and foster innovation. This, in turn, diminishes the EU's competitive standing on the global stage. Moreover, Articles 3 and 4 of the CDSM Directive, governing exceptions to copyright infringement during TDM on copyrighted works, come with complexities and restrictions. They create an undesirable distinction between TDM for research and other purposes, adding uncertainty for copyright holders and innovators.

⁴³³ Chiou, "Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?," 411.

⁴³⁴ Meys, "Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market," 473.

⁴³⁵ Margoni and Kretschmer, "A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology," 687.

⁴³⁶ Chiou, "Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?," 411.

⁴³⁷ Chiou, 411.

Copyright in the digital era is a global problem, as illustrated by the different approaches taken by developed countries like the US, Japan, and EU member states. The US enhances the fair use exception to copyright under certain conditions for ML activities. In contrast, Japan demonstrates remarkable flexibility by stating that machines do not enjoy works and are therefore free from copyright restrictions. Alongside these countries, Türkiye, a developing state within the *droit d’auteur* system, has not defined any copyright exceptions and is examined to assess possible infringements that TDM activities may cause. Overall, this study's comparison reveals that the numerous concerns surrounding TDM exceptions pose a risk to innovative TDM developments in the EU, as businesses might relocate to jurisdictions like Japan or the United States that offer more legal certainty for innovation and TDM.⁴³⁸

⁴³⁸ Gerrish and Skavlan, “European Copyright Law and the Text and Data Mining Exceptions and Limitations,” 59.

References

- “ DSM Directive Implementation Tracker.” Accessed January 13, 2024. <https://www.notion.so/DSM-Directive-Implementation-Tracker-361cfac48e814440b353b32692bba879>.
- “Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG).” Accessed January 11, 2024. https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0328.
- “AI Act | Shaping Europe’s Digital Future,” May 29, 2024. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.
- Aksu, Mustafa, Elif Gizem Koç, Sevgi Erarslan, Gülce Erek, Gülden Mehmed, Batuhan Akkuş, and Muhammed Enes Yıldız. *Yapay Zeka ve Hukuk*. Edited by Mustafa Aksu. 1st ed. İstanbul: Onikilevha Yayıncılık, 2023.
- Aldoseri, Abdulaziz, Khalifa N. Al-Khalifa, and Abdel Magid Hamouda. “Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges.” *Applied Sciences* 13, no. 12 (January 2023): 7082. <https://doi.org/10.3390/app13127082>.
- “Article L122-5-3 - Code de La Propriété Intellectuelle - Légifrance.” Accessed January 12, 2024. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044363192#.
- “Artificial Intelligence / The Government of Japan - JapanGov -.” Accessed December 15, 2023. https://www.japan.go.jp/tomodachi/2018/spring2018/artificial_intelligence.html.
- Ateş, Mustafa. *Fikrî Hukukta Eser*. Ankara: Turhan Kitapevi, 2007.
- . “Veri tabanlarının Hukuki Koruması.” *Ankara Üniversitesi Hukuk Fakültesi Dergisi* 55, no. 1 (2006): 1. https://doi.org/10.1501/Hukfak_00000000365.
- “Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).” *U.S. Copyright Office Fair Use Index*, n.d. <http://copyright.gov/fair-use/index.html>.
- Aydıncık, Şirin. *Fikri haklara ilişkin lisans sözleşmeleri*. İstanbul: Arıkan Basım Yayım, 2006.
- Azeroual, Otmane. “Text and Data Quality Mining in CRIS.” *Information* 10, no. 12 (December 2019): 374. <https://doi.org/10.3390/info10120374>.

- Baio, Andy. “AI Data Laundering: How Academic and Nonprofit Researchers Shield Tech Companies from Accountability.” *Waxy.Org* (blog), September 30, 2022. <https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/>.
- . “Invasive Diffusion: How One Unwilling Illustrator Found Herself Turned into an AI Model.” *Waxy.Org* (blog), November 1, 2022. <https://waxy.org/2022/11/invasive-diffusion-how-one-unwilling-illustrator-found-herself-turned-into-an-ai-model/>.
- Bayamlıoğlu, İbrahim Emre. “Fikir ve Sanat Eserleri Hukukunda Teknolojik Koruma.” Phd Thesis, Marmara Üniversitesi, 2007.
- Bernier, Alexander, Christian Busse, and Tania Bubela. “Public Biological Databases and the Sui Generis Database Right.” *IIC - International Review of Intellectual Property and Competition Law* 54, no. 9 (October 1, 2023): 1316–58. <https://doi.org/10.1007/s40319-023-01373-0>.
- Big Think. “How AI Learned to Paint like Rembrandt,” July 26, 2021. <https://bigthink.com/high-culture/ai-learned-paint-like-rembrandt/>.
- Bornmann, Lutz, Robin Haunschild, and Rüdiger Mutz. “Growth Rates of Modern Science: A Latent Piecewise Growth Curve Approach to Model Publication Numbers from Established and New Literature Databases.” *Humanities and Social Sciences Communications* 8, no. 1 (October 7, 2021): 1–15. <https://doi.org/10.1057/s41599-021-00903-w>.
- Bottis, Maria, Marinos Papadopoulos, Christos Zampakolas, and Paraskevi Ganatsiou. “Text and Data Mining in the EU ‘Acquis Communautaire’ Tinkering with TDM & Digital Legal Deposit.” *Erasmus Law Review* 12, no. 2 (2019): 190–208. <https://doi.org/10.5553/ELR.000140>.
- Bundesministerium der Justiz. “Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes.” Accessed January 11, 2024. https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/DE/2020_Gesetz_Anpassung-Urheberrecht-dig-Binnenmarkt.html.
- Cadwalladr, Carole, and Emma Graham-Harrison. “Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.” *The Guardian*, March 17, 2018, sec. News. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

- Cahan, Eli M., Tina Hernandez-Boussard, Sonoo Thadaney-Israni, and Daniel L. Rubin. "Putting the Data before the Algorithm in Big Data Addressing Personalized Healthcare." *Npj Digital Medicine* 2, no. 1 (August 19, 2019): 1–6. <https://doi.org/10.1038/s41746-019-0157-2>.
- Cano, Jenn. "The V's of Big Data Velocity Volume Value Variety and Veracity." Accessed June 17, 2024. <https://www.coursehero.com/file/65841733/The-Vs-of-Big-Data-Velocity-Volume-Value-Variety-and-Veracitypdf/>.
- Carroll, Michael W. "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful." *UC Davis Law Review- American University*, WCL Research Paper No. 2020-15, 53 (December 1, 2019): 893.
- Castillo, Dianne. "Generative AI vs. Machine Learning." Seldon, November 10, 2023. <https://www.seldon.io/generative-ai-vs-machine-learning>.
- CBC Radio*. "Can an Artist Sue an AI over Copyright Infringement?" October 12, 2018. <https://www.cbc.ca/radio/spark/409-1.4860495/can-an-artist-sue-an-ai-over-copyright-infringement-1.4860762>.
- "CDSM Implementation Resource Page – CREATE." Accessed January 12, 2024. <https://www.create.ac.uk/cdsm-implementation-resource-page/#consultations-transpositions>.
- Çelik, Abdullah. "Fikir ve Sanat Eserleri Üzerindeki Çoğaltma ve Yayma Haklarının İhlali, İhlalin Sonuçları." Phd Thesis, Dokuz Eylül Üniversitesi, 2010. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>.
- Chang, Yupeng, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, et al. "A Survey on Evaluation of Large Language Models." arXiv, October 17, 2023. <http://arxiv.org/abs/2307.03109>.
- Cheshmeh Sohrabi, Mehrdad, and Amir Mashhadi. "Using Data Mining, Text Mining, and Bibliometric Techniques to the Research Trends and Gaps in the Field of Language and Linguistics." *Journal of Psycholinguistic Research* 52, no. 2 (April 1, 2023): 607–30. <https://doi.org/10.1007/s10936-022-09911-6>.
- "China and the Great Global AI Governance Divide | Lowy Institute." Accessed June 12, 2024. <https://www.lowyinstitute.org/the-interpreter/china-great-global-ai-governance-divide>.
- Chiou, Theodoros. "Copyright Lessons on Machine Learning: What Impact on Algorithmic Art?," 2019. https://www.jipitec.eu/issues/jipitec-10-3-2019/5025/chiou_pdf.pdf.

- Christensen, Kristina. “A European Solution for Text and Data Mining in the Development of Creative Artificial Intelligence” 4, no. 2 (2021).
- Christie, Andrew F., and Robin Wright. “A Comparative Analysis of the Three-Step Tests in International Treaties.” *IIC - International Review of Intellectual Property and Competition Law* 45, no. 4 (June 2014): 409–33. <https://doi.org/10.1007/s40319-014-0202-2>.
- Commission Staff Working Document Türkiye 2023 Report Accompanying the document Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 2023 Communication on EU Enlargement policy (2023). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=SWD:2023:696:FIN>.
- COMMUNIA Association. “Policy Paper #15 on Using Copyrighted Works for Teaching the Machine.” Accessed January 15, 2024. <https://communia-association.org/policy-paper/policy-paper-15-on-using-copyrighted-works-for-teaching-the-machine/>.
- Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions- Making the most of the EU’s innovative potential An intellectual property action plan to support the EU’s recovery and resilience (2020). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020DC0760>.
- Çonkar, Mustafa Halil. “Dijital Tek Pazarda Telif Hakkı ve Bağlantılı Haklar.” In *Türk hukukunun Avrupa Birliği hukukuna uyumu özel hukuk: acquis communautaire’in alınması: açıklamalar, değerlendirmeler, öneriler = Approximation of Turkish law with European Union law private law: transposition of acquis communautaire: remarks, assesments, suggestions*, edited by Arslan Kaya, Baki İlkay Engin, Ali Pashlı, Şehriban İpek Aşıkoğlu, and Elif Oğuz. İstanbul Üniversitesi yayınları, no 5269. İstanbul: İstanbul Üniversitesi Hukuk Fakültesi, 2020.
- “Copyright Law of Japan | Copyright Law of Japan | Copyright Research and Information Center CRIC.” Accessed December 15, 2023. <https://www.cric.or.jp/english/clj/cl2.html>.
- Court of Justice of the European Union, Opinion of Advocate General Trstenjak delivered on 12 February 2009. Infopaq International A/S v Danske

- Dagblades Forening. Reference for a preliminary ruling: Højesteret - Denmark. Case C-5/08. (ECJ 2009).
- Court of Justice of the European Union, Order of the Court (Third Chamber), 17 January 2012. Infopaq International A/S v Danske Dagblades Forening. Case C-302/10. (ECJ 2012).
- Craig, Carys J. “Globalizing User Rights-Talk: On Copyright Limits and Rhetorical Risks.” *American University International Law Review* Vol 33, no. No. 1 (December 2, 2017). <https://ssrn.com/abstract=3109777>.
- CV Online Latvia v Melons, No. C-762/19, ECLI:EU:C:2021:434. (Court of Justice of the European Union June 3, 2021).
- “Data Act: Council Adopts New Law on Fair Access to and Use of Data.” Accessed April 6, 2024. <https://www.consilium.europa.eu/en/press/press-releases/2023/11/27/data-act-council-adopts-new-law-on-fair-access-to-and-use-of-data/>.
- “Data for Sale | YaleGlobal Online.” Accessed November 30, 2023. <https://archive-yaleglobal.yale.edu/content/data-sale>.
- Deb, Jayeesha. “AI Art Generates Controversy Among Artists.” *The Arcadia Quill* (blog). Accessed June 12, 2024. <https://arcadiaquill.com/25242/showcase/ai-art-generates-controversy-among-artists/>.
- Derclaye, Estelle. “The Court of Justice Copyright Case Law: Quo Vadis?” 36, no. 11 (November 1, 2014). <https://nottingham-repository.worktribe.com/index.php/output/14319944/the-court-of-justice-copyright-case-law-quo-vadis>.
- Derclaye, Estelle, and Martin Husovec. “Sui Generis Database Protection 2.0: Judicial and Legislative Reforms.” *European Intellectual Property Review (EIPR)* - *Forthcoming*, November 16, 2021. Available at SSRN: <https://ssrn.com/abstract=3964943> or <http://dx.doi.org/10.2139/ssrn.3964943>.
- Dermawan, Artha. “Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn from the Japanese ‘Nonenjoyment’ Purposes?” *The Journal of World Intellectual Property*, May 27, 2023, jwip.12285. <https://doi.org/10.1111/jwip.12285>.
- “DG Internal Market and Services Working Paper: First Evaluation of Directive 96/9/EC on the Legal Protection of Databases.” Brussels: European Commission, December 12, 2005. <https://openfuture.eu/wp->

- content/uploads/2021/12/2019EC-evaluation_report_legal_protection_databases_december_2005_en.pdf.
- “Directive - 2019/790 - EN - Dsm - EUR-Lex.” Accessed April 23, 2024. <https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX%3A32019L0790>.
- Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, 077 OJ L § (1996). <http://data.europa.eu/eli/dir/1996/9/oj/eng>.
- Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, 167 OJ L § (2001). <http://data.europa.eu/eli/dir/2001/29/oj/eng>.
- Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L130/92 Copyright and related rights in the Digital Single Market § (2019).
- Doctorow, Cory, and Katharine Trendacosta. “AI Art Generators and the Online Image Market.” Electronic Frontier Foundation, April 3, 2023. <https://www.eff.org/deeplinks/2023/04/ai-art-generators-and-online-image-market>.
- Dorokhin, Victor S. “The Fair Use Doctrine in U.S. Copyright Law.” *Medium* (blog), October 16, 2023. <https://medium.com/@victordorokhin1/the-fair-use-doctrine-in-u-s-copyright-law-e65c98049946>.
- Ducato, Rossana, and Alain Strowel. “Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out.” *European Intellectual Property Review*, CRIDES Working Paper Series no. 1/2021, February 1, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829858.
- . “Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to ‘Machine Legibility.’” *IIC - International Review of Intellectual Property and Competition Law* 50, no. 6 (July 2019): 649–84. <https://doi.org/10.1007/s40319-019-00833-w>.
- Duffin, Matthew. “Machine Unlearning: The Critical Art of Teaching AI to Forget.” *VentureBeat* (blog), August 12, 2023. <https://venturebeat.com/ai/machine-unlearning-the-critical-art-of-teaching-ai-to-forget/>.

- Dusollier, Severine. “The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition.” *Common Market Law Review* 57, no. Issue 4 (August 1, 2020): 979–1030. <https://doi.org/10.54648/COLA2020713>.
- Edwards, Benj. “University of Chicago Researchers Seek to ‘Poison’ AI Art Generators with Nightshade.” *Ars Technica*, October 25, 2023. <https://arstechnica.com/information-technology/2023/10/university-of-chicago-researchers-seek-to-poison-ai-art-generators-with-nightshade/>.
- Eechoud, Mireille van. “Along the Road to Uniformity – Diverse Readings of the Court of Justice Judgments on Copyright Work,” 2012.
- Ercan, Saniye Ezgi. “Avrupa Birliği Dijital Tek Pazarda Telif Hakkı Yönergesi ile Öngörülen İstisnalar Çerçevesinde Türk Fikir ve Sanat Eserleri Kanununun Değerlendirilmesi.” Masters Thesis, İhsan Doğramacı Bilkent Üniversitesi, 2022. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>.
- Espinoza, Javier. “European Companies Sound Alarm over Draft AI Law.” *Financial Times*, June 30, 2023, sec. EU tech regulation. <https://www.ft.com/content/9b72a5f4-a6d8-41aa-95b8-c75f0bc92465>.
- European Commission. “Copyright: Commission Calls on Member States to Comply with EU Rules on Copyright in the Digital Single Market.” Text. Accessed June 6, 2024. https://ec.europa.eu/commission/presscorner/detail/en/MEX_21_3902.
- European Commission. Impact Assessment on the modernisation of EU copyright rules, Pub. L. No. SWD(2016) 301 (2016). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52016SC0301>.
- . “White Paper on Artificial Intelligence – A European Approach to Excellence and Trust.” Brussels, 2020. https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.
- European Commission - European Commission. “11 Member States Referred to the Court of Justice of the EU.” Text. Accessed January 10, 2024. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_704.
- European Commission - European Commission. “Copyright: MSs Urged to Fully Transpose EU Copyright Rules.” Text. Accessed June 6, 2024. https://ec.europa.eu/commission/presscorner/detail/EN/IP_22_2692.

- Fiil-Flynn, Sean M., Brandon Butler, Michael Carroll, Or Cohen-Sasson, Carys Craig, Lucie Guibault, Peter Jaszi, et al. "Legal Reform to Enhance Global Text and Data Mining Research." *Science* 378, no. 6623 (December 2, 2022): 951–53. <https://doi.org/10.1126/science.add6124>.
- Fikir ve Sanat Eserleri Kanunu, Pub. L. No. Yayınlandığı R.Gazete : Tarih : 13/12/1951 Sayı : 7981 Yayınlandığı Düstur : Tertip : 3 Cilt: 33 Sayfa : 49, 5846 (1951). <https://www.mevzuat.gov.tr/mevzuatmetin/1.3.5846.pdf>.
- Filippov, Sergey, and Paul Hofheinz. "Text and Data Mining for Research and Innovation." *The Lisbon Council*, no. Issue 20/2016 (2016). https://lisboncouncil.net/wp-content/uploads/2020/08/LISBON_COUNCIL_Text_and_Data_Mining_for_Research_and_Innovation.pdf.
- Flynn, Sean, Christophe Geiger, João Quintais, Thomas Margoni, Matthew Sag, L. Guibault, and Michael W. Carroll. "Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action." *SSRN Electronic Journal*, 2020. <https://doi.org/10.2139/ssrn.3578819>.
- Frosio, Giancarlo F. "Communia Final Report on the Digital Public Domain," March 31, 2011. https://www.communia-association.org/wp-content/uploads/COMMUNIA_Final_Report.pdf.
- Fukuyama, Mayumi. "Society 5.0: Aiming for a New Human-Centered Society." *Japan Spotlight*, August 2018, 47–50.
- Geiger, Christophe, Giancarlo Frosio, and Oleksandr Bulayenko. "Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/Eu." *SSRN Electronic Journal*, 2019. <https://doi.org/10.2139/ssrn.3470653>.
- . "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects." *SSRN Electronic Journal*, 2018. <https://doi.org/10.2139/ssrn.3160586>.
- Gerrish, Charlotte, and Anders Molander Skavlan. "European Copyright Law and the Text and Data Mining Exceptions and Limitations" 2, no. 2 (2019).
- Gervais, Daniel J. "Exploring the Interfaces Between Big Data and Intellectual Property Law." *SSRN Electronic Journal*, 2019. <https://doi.org/10.2139/ssrn.3360344>.
- Güçlütürk, Osman Gazi. "Türk Hukukunda Makine Öğrenmesine Dayalı Yapay Zekada Verinin Hukuka Uygun Şekilde Kullanılması." Phd Thesis,

Galatasaray Üniversitesi, 2021.
<https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=DUkPWcgZfXOGAXLTXfFIw&no=dEVj8z85tRmLg-yXXWxLtA>.

Güleş, Tuğba. *Technological Advances in Plant Innovation and the Dynamics of Intellectual Property Law*. Birinci baskı. Seçkin. Hukuk, no: 1195. Ankara: Seçkin, 2018.

Gupta, Maanak, Charankumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy." *IEEE Access* 11 (2023): 80218–45. <https://doi.org/10.1109/ACCESS.2023.3300381>.

Hacker, Philipp. "A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act." *Law, Innovation and Technology* 13, no. 2 (July 3, 2021): 257–301. <https://doi.org/10.1080/17579961.2021.1977219>.

Handke, Christian, Lucie Guibault, and Joan-Josep Vallbb. "Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research." *SSRN Electronic Journal*, 2015. <https://doi.org/10.2139/ssrn.2608513>.

Haojing Chen, YP Tsang, and CH Wu. "When Text Mining Meets Science Mapping in the Bibliometric Analysis: A Review and Future Opportunities." *International Journal of Engineering Business Management* Volume 15 (December 2023). <https://doi.org/doi.org/10.1177/18479790231222349>.

Helvacı, Serap, and Eray Aksın Atar. "Fikir ve Sanat Eserleri Hukukunda Üç Aşamalı Test." *Marmara Üniversitesi Hukuk Fakültesi Hukuk Araştırmaları Dergisi* 22, no. 3 (December 16, 2016): 1289–1302.

Henriksen, Ellen Emilie. "Big Data, Microtargeting, and Governmentality in Cyber-Times. The Case of the Facebook-Cambridge Analytica Data Scandal." Master thesis, University of Oslo, 2019. <http://www.duo.uio.no/>.

Henshall, Will. "E.U.'s AI Regulation Could Be Softened After Pushback." *TIME*, November 22, 2023. <https://time.com/6338602/eu-ai-regulation-foundation-models/>.

Hervey, Matt. "The EU AI Act and IP." *LoupedIn*, June 26, 2023. <https://loupedin.blog/2023/06/the-eu-ai-act-and-ip/>.

Hilty, Reto M, Christophe Geiger, and Jonathan Griffiths. "Towards a Balanced Interpretation of the 'Three-Step Test' in Copyright Law." *European Intellectual Property Review*, no. Vol.4 (2016): 489-496.

- “Hiroshima Process International Guiding Principles for Advanced AI System | Shaping Europe’s Digital Future,” October 30, 2023. <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system>.
- House, The White. “FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.” The White House, October 30, 2023. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>.
- Hugenholtz, P. Bernt. “The New Copyright Directive: Text and Data Mining (Articles 3 and 4).” Kluwer Copyright Blog, July 24, 2019. <https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>.
- Hugenholtz, P. Bernt, and Ruth Okediji. “Conceiving an International Instrument on Limitations and Exceptions to Copyright.” *SSRN Electronic Journal*, 2012. <https://doi.org/10.2139/ssrn.2017629>.
- Hugenholtz, P. Bernt, and Martin Kretschmer. *Copyright Reconstructed: Rethinking Copyright’s Economic Rights in a Time of Highly Dynamic Technological and Economic Change Chapter 1: Reconstructing Rights: Project Synthesis and Recommendations*. Information Law Series, volume 41. Alphen aan den Rijn, The Netherlands: Wolters Kluwer, 2018.
- “Infopaq International A/S v. Danske Dagblades Forening (Case C-5/08), [2012] BUS LR 102 – Center for Study and Research in Intellectual Property Rights [CSRIPR], NUSRL.” Accessed January 18, 2024. <https://csripnusrl.wordpress.com/2022/04/13/infopaq-international-a-s-v-danske-dagblades-forening-case-c-5-08-2012-bus-lr-102/>.
- Intuji. “How Does Google Search Work? The Google Algorithm Explained,” February 3, 2023. <https://intuji.com/how-does-google-search-work/>.
- Justia Law. “Authors Guild, Inc. v. HathiTrust, No. 12-4547 (2d Cir. 2014).” Accessed December 16, 2023. <https://law.justia.com/cases/federal/appellate-courts/ca2/12-4547/12-4547-2014-06-10.html>.
- Kaplan, Menachem, Marissa Yu, and Luc Von Danwitz. “The Fair Use Defense for Generative AI Tools After Warhol V. Goldsmith.” Passle, June 16, 2023.

<https://blog.freshfields.us//post/102ih2k/the-fair-use-defense-for-generative-ai-tools-after-warhol-v-goldsmith>.

Karaaslan, Pelin. “Fikir ve Sanat Eserlerinin Teknik Gereklik Nedeniyle Geçici Çoğaltılması.” *Ticaret ve Fikri Mülkiyet Hukuku Dergisi* C:8, no. S:1 (2022): 29–49.

Karanikolova, Kristina, Julien Chicot, Anna Gkogka, Robbert Fisher, Lionel Bently, Gabriela Bodea, Estelle Derclaye, et al. *Study in Support of the Evaluation of Directive 96/9/EC on the Legal Protection of Databases: Final Report*. Publications Office of the European Union, 2018. <https://data.europa.eu/doi/10.2759/04895>.

Kariri, Elham, Hassen Louati, Ali Louati, and Fatma Masmoudi. “Exploring the Advancements and Future Research Directions of Artificial Neural Networks: A Text Mining Approach.” *Applied Sciences* 13, no. 5 (January 2023): 3186. <https://doi.org/10.3390/app13053186>.

Karpathy, Andrej, Pieter Abeel, Greg Brockman, Peter Chen, Vicki Cheung, Yan Duan, Ian Goodfellow, et al. “Generative Models.” OpenAI.com, June 16, 2016. <https://openai.com/research/generative-models>.

Keller, Paul. “DSM Implementation: What Is Happening with Regards to the Other Provisions,” February 15, 2021. <https://eu-copyright-implementation.info/en/comments-on-implementation/dsm-implementation-what-is-happening-with-regards-to-the-other-provisions/>.

———. “Generative AI and Copyright: Convergence of Opt-Outs?” Kluwer Copyright Blog, November 23, 2023. <https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/>.

———. “Protecting Creatives or Impeding Progress? Machine Learning and the EU Copyright Framework.” Kluwer Copyright Blog, February 20, 2023. <https://copyrightblog.kluweriplaw.com/2023/02/20/protecting-creatives-or-impeding-progress-machine-learning-and-the-eu-copyright-framework/>.

Keller, Paul, and Zuzanna Warso. “Defining Best Practices for Opting out of ML Training.” Open Future EU, September 29, 2023. https://openfuture.eu/wp-content/uploads/2023/09/Best-_practices_for_optout_ML_training.pdf.

Kleinsmith, Erik. “Combating Fake News with Data Mining.” *Edge* (blog), July 22, 2019. <https://amuedge.com/combating-fake-news-with-data-mining/>.

- Klosowski, Thorin. “No Robots(.Txt): How to Ask ChatGPT and Google Bard to Not Use Your Website for Training.” Electronic Frontier Foundation, December 12, 2023. <https://www.eff.org/deeplinks/2023/12/no-robotstxt-how-ask-chatgpt-and-google-bard-not-use-your-website-training>.
- Knibbs, Kate. “A New Tool Helps Artists Thwart AI—With a Middle Finger.” *Wired*. Accessed January 15, 2024. <https://www.wired.com/story/kudurru-ai-scraping-block-poisoning-spawning/>.
- . “Most Top News Sites Block AI Bots. Right-Wing Media Welcomes Them.” *Wired*. Accessed April 8, 2024. <https://www.wired.com/story/most-news-sites-block-ai-bots-right-wing-media-welcomes-them/>.
- Kollár, Péter. “Mind If I Mine? A Study on the Justification and Sufficiency of Text and Data Mining Exceptions in the European Union.” LLM Thesis, Maastricht university, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3960570.
- Koutras, Nikos, and Haydn Rigby. “A Scientific Analysis of the Three-Step Test: Through the Lenses of International and Australian Laws.” *Publishing Research Quarterly* 38, no. 3 (September 2022): 503–18. <https://doi.org/10.1007/s12109-022-09898-x>.
- Kretschmer, Martin, and Thomas Margoni. “Data Mining: Why the EU’s Proposed Copyright Measures Get It Wrong.” *The Conversation*, May 24, 2018. <http://theconversation.com/data-mining-why-the-eus-proposed-copyright-measures-get-it-wrong-96743>.
- Külahcı, İbrahim Hulusi. “Elektronik Veri Tabanlarının Fikir ve Sanat Hukuku Kapsamında Korunması.” Masters Thesis, İstanbul Sabahattin Zaim Üniversitesi, 2019. <https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=vjszP7PzV0HebcjFEvDfwICZTYGOQdmTgR9qC193FdODp4hXqNt2lJRfGxIMUjce>.
- Lazarova, Ana. “The Last in Line: Bulgaria Implements the CDSM Directive.” *Kluwer Copyright Blog*, December 27, 2023. <https://copyrightblog.kluweriplaw.com/2023/12/27/the-last-in-line-bulgaria-implements-the-cdsm-directive/>.
- Lemley, Mark A., and Bryan Casey. “Fair Learning.” *SSRN Electronic Journal*, 2020. <https://doi.org/10.2139/ssrn.3528447>.

- Levent Yavuz, Türkay Alica, and Fethi Merdivan. *Fikir ve Sanat Eserleri Kanunu Yorumu*. 2.Baskı. Vol. I. 2 vols. Ankara: Seçkin Yayıncılık, 2014.
- . *Fikir ve Sanat Eserleri Kanunu Yorumu*. 2.baskı. Vol. II. 2 vols. Ankara: Seçkin Yayıncılık, 2014.
- LII / Legal Information Institute. “17 U.S. Code § 107 - Limitations on Exclusive Rights: Fair Use.” Accessed December 16, 2023. <https://www.law.cornell.edu/uscode/text/17/107>.
- Malharbe, Aude. “L’évolution du droit d’auteur face aux défis des intelligences artificielles génératives.” Conseil National Des Barreaux. Accessed January 12, 2024. <https://consultation.avocat.fr/blog/aude-malherbe/article-170740-l-evolution-du-droit-d-auteur-face-aux-defis-des-intelligences-artificielles-generatives.html>.
- Margoni, Thomas. “Computational Legal Methods: Text and Data Mining in Intellectual Property Research.” In *Handbook of Intellectual Property Research*, edited by Irene Calboli and Maria Lillà Montagnani, 1st ed., 487-C31.P47. Oxford University PressOxford, 2021. <https://doi.org/10.1093/oso/9780198826743.003.0032>.
- Margoni, Thomas, Thomas Gils, and Eyüp Kun. “Chapter X of the Data Act and the Sui Generis Database Right.” CiTiP blog, June 14, 2022. <https://www.law.kuleuven.be/citip/blog/chapter-10-of-the-data-act-and-the-sui-generis-database-right/>.
- Margoni, Thomas, and Martin Kretschmer. “A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology.” *GRUR International* 71, no. 8 (2022): 685–701. <https://doi.org/10.1093/grurint/ikac054>.
- Maslej, Nestor, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, et al. “The AI Index 2023 Annual Report.” Stanford, CA: Institute for Human-Centered AI, Stanford University, April 2023. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf.
- Mat. “Spawning Opts out 78 Million Artworks from AI Training,” May 30, 2023. <https://spawning.substack.com/p/spawning-opts-out-78-million-artworks>.
- “Mevzuat Bilgi Sistemi.” Accessed January 16, 2024. <https://mevzuat.tbmm.gov.tr/>.

- Meys, Romain. “Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?” *GRUR International* 69, no. 5 (May 1, 2020): 457–73. <https://doi.org/10.1093/grurint/ikaa046>.
- Miranda Branco Tomé Quintais, João Pedro de. “Copyright in the Age of Online Access: Alternative Compensation Systems in EU Copyright Law.” Phd Thesis, Universiteit van Amsterdam, 2017. https://pure.uva.nl/ws/files/7607912/QUINTAIS_Doctoral_Dissertation_Final_Manuscript.pdf.
- “Mission du CSPLA sur les exceptions de fouille de textes et de données (‘text and data mining’).” Accessed January 17, 2024. <https://www.culture.gouv.fr/Nous-connaître/Organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA/Mission-du-CSPLA-sur-les-exceptions-de-fouille-de-textes-et-de-données-text-and-data-mining>.
- MIT News | Massachusetts Institute of Technology. “Explained: Neural Networks,” April 14, 2017. <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.
- MIT Technology Review. “Generative AI Is Changing Everything. But What’s Left When the Hype Is Gone?” Accessed January 15, 2024. <https://www.technologyreview.com/2022/12/16/1065005/generative-ai-revolution-art/>.
- Mittermaier, Mirja, Mariam M. Raza, and Joseph C. Kvedar. “Bias in AI-Based Models for Medical Applications: Challenges and Mitigation Strategies.” *Npj Digital Medicine* 6, no. 1 (June 14, 2023): 1–3. <https://doi.org/10.1038/s41746-023-00858-z>.
- Murray-Rust, Peter. “The Right to Read Is the Right to Mine.” Open Knowledge Foundation blog, June 1, 2012. <https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/>.
- Nordemann, Jan Bernd, Jonathan Pukas, and Julian Waiblinger. “The EU DSM Copyright Directive: Implementation in Germany 2021 – Part I.” Kluwer Copyright Blog, January 16, 2023. <https://copyrightblog.kluweriplaw.com/2023/01/16/the-eu-dsm-copyright-directive-implementation-in-germany-2021-part-i/>.

- “Number of Academic Papers Published Per Year – WordsRated.” Accessed January 18, 2024. <https://wordrated.com/number-of-academic-papers-published-per-year/#>.
- OECD. “Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by ‘Big Data.’” OECD Digital Economy Papers. Vol. 222. OECD Digital Economy Papers, June 18, 2013. <https://doi.org/10.1787/5k47zw3fcp43-en>.
- Okazaki, Shintaro, Ana M. Díaz-Martín, Mercedes Rozano, and Héctor David Menéndez-Benito. “Using Twitter to Engage with Customers: A Data Mining Approach.” *Internet Research* 25, no. 3 (June 1, 2015): 416–34. <https://doi.org/10.1108/IntR-11-2013-0249>.
- Open Future. “A First Look at the Copyright Relevant Parts in the Final AI Act Compromise.” Accessed December 24, 2023. <https://openfuture.eu/blog/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise>.
- Open Letter. “Open Letter to the Representatives of the European Commission, the European Council and the European Parliament.” Accessed December 25, 2023. <https://www.igizmo.it/wp-content/uploads/2023/06/Open-Letter-EU-AI-Act-and-Signatories.pdf>.
- “Ouvrir La Science - La Fouille de Textes et de Données à Des Fins de Recherche : Une Pratique Confirmée et Désormais Opérationnelle En Droit Français.” Accessed January 13, 2024. <https://www.ouvrirlascience.fr/la-fouille-de-textes-et-de-donnees-a-des-fins-de-recherche-une-pratique-confirmee-et-desormais-operationnelle-en-droit-francais/>.
- Özçelik, Ş.Bariş. “Yapay Zekanın Veri Koruma, Sorumluluk ve Fikri Mülkiyet Açısından Ortaya Çıkardığı Hukuki Gereksinimler.” *Adalet Dergisi*, no. 66 (May 2021): 87–116.
- Özdemir, Süleyman Emre. “Eseri Çoğaltma Hakkının Cezai Himayesi ve Şahsen Kullanma Meselesi.” *Marmara Üniversitesi Hukuk Fakültesi Hukuk Araştırmaları Dergisi* 28, no. 1 (June 30, 2022): 170–83. <https://doi.org/10.33433/maruhad.1094016>.
- Pain, Elisabeth. “How to Keep up with the Scientific Literature | Science | AAAS.” Accessed December 16, 2023. <https://www.science.org/content/article/how-keep-scientific-literature>.

- Peres, Renana, Martin Schreier, David Schweidel, and Alina Sorescu. “On ChatGPT and beyond: How Generative Artificial Intelligence May Affect Research, Teaching, and Practice.” *International Journal of Research in Marketing* 40, no. 2 (June 2023): 269–75. <https://doi.org/10.1016/j.ijresmar.2023.03.001>.
- Placido, Dani Di. “The Problem With AI-Generated Art, Explained.” *Forbes*. Accessed June 12, 2024. <https://www.forbes.com/sites/danidiplacido/2023/12/30/ai-generated-art-was-a-mistake-and-heres-why/>.
- Popescu, Andrei-Dragoș. “The Value of Data From an Artificial Intelligence Perspective.” *Annals of the University of Craiova for Journalism, Communication and Management* Volume 5 (November 2019): 172–94.
- Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (2022). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2022:68:FIN>.
- Quintais, João Pedro. “Generative AI, Copyright and the AI Act.” *Kluwer Copyright Blog*, May 9, 2023. <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/>.
- . “The New Copyright in the Digital Single Market Directive: A Critical Look.” *SSRN Electronic Journal*, 2019. <https://doi.org/10.2139/ssrn.3424770>.
- Recital 105 | EU Artificial Intelligence Act. Accessed June 12, 2024. <https://artificialintelligenceact.eu/recital/105/>.
- Riedl, Mark. “A Very Gentle Introduction to Large Language Models without the Hype.” *Medium* (blog), May 25, 2023. <https://mark-riedl.medium.com/a-very-gentle-introduction-to-large-language-models-without-the-hype-5f67941fa59e>.
- Rosati, Eleonora. “An EU Text and Data Mining Exception for the Few: Would It Make Sense?” *Journal of Intellectual Property Law & Practice* 13, no. 6 (June 1, 2018): 429–30. <https://doi.org/10.1093/jiplp/jpy063>.
- . “Copyright at the CJEU: Back to the Start (of Copyright Protection).” In *Developments and Directions in Intellectual Property Law*, edited by Hayleigh Bosher and Eleonora Rosati, 1st ed., 211-C13P83. Oxford University PressOxford, 2023. <https://doi.org/10.1093/oso/9780192864475.003.0013>.

- . *Handbook on the Economics of Copyright: A Guide for Students and Teachers*. Edited by Richard Watt. Edward Elgar Publishing, 2014.
- . “The DSM Directive Two Years On: Do Things Ever Get Easier?” *Iic; International Review of Industrial Property and Copyright Law* 52, no. 9 (2021): 1139. <https://doi.org/10.1007/s40319-021-01082-6>.
- . “The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects.” *Briefing Requested by the JURI Committee, Policy Department for Citizens’ Rights and Constitutional Affairs, European Parliament*, n.d.
- “Sacem, in Favour of Virtuous, Transparent, and Fair AI, Exercises Its Right to Opt-out - SACEM.” Accessed January 13, 2024. <https://societe.sacem.fr/en/news/our-society/sacem-favour-virtuous-transparent-and-fair-ai-exercises-its-right-opt-out>.
- Sag, Matthew. “The New Legal Landscape for Text Mining and Machine Learning.” *Journal of the Copyright Society of the USA* Vol. 66 (February 27, 2020): p.291. <http://dx.doi.org/10.2139/ssrn.3331606>.
- Samuelson, Pamela. “Questioning a New Intellectual Property Right for Press Publishers – Communications of the ACM,” March 1, 2019. <https://cacm.acm.org/opinion/questioning-a-new-intellectual-property-right-for-press-publishers/>.
- . “Regulating Technology Through Copyright Law: A Comparative Perspective.” *European Intellectual Property Review*, no. 42 (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635094.
- . “The Digital Dilemma: A Perspective on Intellectual Property in the Information Age,” n.d. <https://people.ischool.berkeley.edu/~pam/papers/digdilsyn.pdf>.
- Şener, Yavuz Selim. “Fikri Mülkiyet Hukukunda Dijital Veri Tabanlarının Korunması.” Masters Thesis, İstanbul Kültür Üniversitesi, 2013. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>.
- Senfleben, Martin. “Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue?” *IIC - International Review of Intellectual Property and Competition Law* 53, no. 10 (November 2022): 1477–1505. <https://doi.org/10.1007/s40319-022-01266-8>.

- . “Copyright, Limitations and the Three-Step Test. An Analysis of the Three-Step Test in International and EC Copyright Law.” Phd Thesis, Universiteit van Amsterdam, 2004.
https://pure.uva.nl/ws/files/3645247/37378_Thesis.pdf.
- . “Generative AI and Author Remuneration.” *IIC - International Review of Intellectual Property and Competition Law* 54, no. 10 (November 1, 2023): 1535–60. <https://doi.org/10.1007/s40319-023-01399-4>.
- Senftleben, Martin, Thomas Margoni, Daniel Antal, Balázs Bodó, Stef van Gompel, Christian Handke, Joost Poort, João Quintais, and Sebastian Schwemer. “Ensuring the Visibility and Accessibility of European Creative Content on the World Market.” *Journal of Intellectual Property, Information Technology and E-Commerce Law* Vol. 13, no. No.1 (2022): 67–86.
<http://dx.doi.org/10.2139/ssrn.3785272>.
- Shaik, Thanveer, Xiaohui Tao, Haoran Xie, Lin Li, Xiaofeng Zhu, and Qing Li. “Exploring the Landscape of Machine Unlearning: A Comprehensive Survey and Taxonomy.” arXiv, October 23, 2023. <http://arxiv.org/abs/2305.06360>.
- “Sony World Photography Award 2023: Winner Refuses Award after Revealing AI Creation.” April 17, 2023. <https://www.bbc.com/news/entertainment-arts-65296763>.
- Spawning AI. “Datadiligence Respect Generative AI Opt-Outs in Your ML Training Pipeline.” Spawning AI. Accessed January 15, 2024.
<https://datadiligence.readthedocs.io/en/latest/readme.html>.
- Suluk, Cahit. “Avrupa Birliği Hukukunda Basın Yayıncılara Tanınan Bağlantılı Haklar.” *İstanbul Medeniyet Üniversitesi Hukuk Fakültesi Dergisi* 7, no. 12 (2022): 169–227.
- Suluk, Cahit, Rauf Karasu, and Temel Nal. *Fikri Mülkiyet Hukuku*. 7th ed. Ankara: Seçkin Yayıncılık, 2023.
- Surden, Harry. “Artificial Intelligence and Law: An Overview.” *Georgia State University Law Review* 35, no. 4 (2019).
<https://scholar.law.colorado.edu/cgi/viewcontent.cgi?article=2340&context=faculty-articles>.
- Synodinou, Tatiana. “Search Engines and Databases in the Search for a Balance: The AG’s Opinion in the ‘CV-Online Latvia’ Case.” Kluwer Copyright Blog, February 22, 2021.

<https://copyrightblog.kluweriplaw.com/2021/02/22/search-engines-and-databases-in-the-search-for-a-balance-the-ags-opinion-in-the-cv-online-latvia-case/>.

Tar, Julia. “Several French Media Block OpenAI’s GPTBot over Data Collection Concerns.” www.euractiv.com, August 29, 2023. <https://www.euractiv.com/section/artificial-intelligence/news/several-french-media-block-openais-gptbot-over-data-collection-concerns/>.

T.C. Yargıtay Başkanlığı. “Yargıtay - Yargıtay İçtihat Merkezi Kullanıma Açılmıştır.” Accessed June 24, 2024. <https://www.yargitay.gov.tr/item/1763/yargitay-ictihat-merkezi-kullanima-acilmistir>.

Team, IBM Data and AI. “AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?” *IBM Blog* (blog), July 6, 2023. <https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/>.

The Art Newspaper - International art news and events. “The Art World’s AI Dilemma: Informed Insight from Industry Experts,” May 30, 2024. <https://www.theartnewspaper.com/2024/05/30/the-art-worlds-ai-dilemma-informed-insight-from-industry-experts>.

“The Copyright Directive Is Failing (Open Letter to Members of the European Parliament and the Council of the European Union).” Accessed December 10, 2023. https://www.create.ac.uk/wp-content/uploads/2018/04/OpenLetter_EU_Copyright_Research_Centres_26_04_2018.pdf.

The Digital Dilemma: Intellectual Property in the Information Age. Accessed June 13, 2024. <https://doi.org/10.17226/9601>.

The Economist. “The Fiction That Turkey Is a Candidate to Join the EU Is Unravelling.” Accessed January 16, 2024. https://www.economist.com/europe/2021/08/26/the-fiction-that-turkey-is-a-candidate-to-join-the-eu-is-unravelling?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18151738051&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAiA75itBhA6EiwAkho9

e-X2cjEaW-PBzfCdEIXWVFHQE6ywQNJ-

ccOVGWgPNUWX3rXaKAVb_RoCJTAQAvD_BwE&gclsrc=aw.ds.

The Economist. “The World Wants to Regulate AI, but Does Not Quite Know How.”

Accessed December 23, 2023.

<https://www.economist.com/business/2023/10/24/the-world-wants-to-regulate-ai-but-does-not-quite-know-how>.

“The European Commission Referred 11 Member States to the Court of Justice of the European Union for Failing to Fully Transpose EU Copyright Rules into National Law | Shaping Europe’s Digital Future,” February 15, 2023. <https://digital-strategy.ec.europa.eu/en/news/european-commission-referred-11-member-states-court-justice-european-union-failing-fully-transpose>.

“The First Piece of AI-Generated Art to Come to Auction.” Accessed June 12, 2024. <https://www.christies.com/en/stories/a-collaboration-between-two-artists-one-human-one-a-machine-0cd01f4e232f4279a525a446d60d4cd1>.

“The Next Rembrandt: Recreating the Work of a Master with AI.” Accessed December 5, 2023. <https://news.microsoft.com/europe/features/next-rembrandt/>.

“Türkiye Büyük Millet Meclisi.” Accessed January 16, 2024. <https://www.tbmm.gov.tr/Yasama/Kanun-Teklifleri-Sonuc>.

Ueno, Tatsuhiro. “The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes – Recent Amendment in Japan and Its Implication.” *GRUR International* 70, no. 2 (March 1, 2021): 145–52. <https://doi.org/10.1093/grurint/ikaa184>.

“Un Décret Précise Les Exceptions En Vue de La Fouille de Textes et de Données.” Accessed January 13, 2024. <https://actualitte.com/article/106646/droit-justice/un-decret-precise-les-exceptions-en-vue-de-la-fouille-de-textes-et-de-donnees>.

Ünal Tekinalp. *Fikri Mülkiyet Hukuku*. 5. Basım. İstanbul: Vedat Kitapçılık, 2012.

“Une Clause-Type Pour s’opposer à La Fouille de Textes et de Données Par Les Intelligences Artificielles - Syndicat National de l’édition.” Accessed January 13, 2024. <https://www.sne.fr/actu/une-clause-type-pour-sopposer-a-la-fouille-de-textes-et-de-donnees-par-les-intelligences-artificielles/>.

“U.S. Copyright Office Fair Use Index.” Accessed December 16, 2023. <https://www.copyright.gov/fair-use/index.html>.

Vincent, James. “The Scary Truth about AI Copyright Is Nobody Knows What Will Happen Next.” *The Verge*, November 15, 2022.

- <https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data>.
- “What Is ChatGPT, DALL-E, and Generative AI? | McKinsey.” Accessed December 13, 2023. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>.
- whitehouse.gov. “Informing Consumers through Smart Disclosure,” March 30, 2012. <https://obamawhitehouse.archives.gov/blog/2012/03/30/informing-consumers-through-smart-disclosure>.
- Wiggers, Kyle. “Spawning Lays out Plans for Letting Creators Opt out of Generative AI Training.” *TechCrunch* (blog), May 3, 2023. <https://techcrunch.com/2023/05/03/spawning-lays-out-its-plans-for-letting-creators-opt-out-of-generative-ai-training/>.
- . “The Current Legal Cases against Generative AI Are Just the Beginning.” *TechCrunch* (blog), January 27, 2023. <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/>.
- WIPO. “WIPO Lex- Law No. 5846 of December 5, 1951, on Intellectual and Artistic Works (as Amended up to Law No. 6552 of September 10, 2014).” Accessed January 16, 2024. <https://www.wipo.int/wipolex/en/text/440051>.
- “WIPO Copyright Treaty,” n.d.
- World Economic Forum January 2011 In Collaboration with Bain & Company, Inc. “Personal Data: The Emergence of a New Asset Class,” January 2011.
- World Intellectual Property Organization, ed. *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)*. WIPO Publication, no. 615. Geneva: World Intellectual Property Organization, 1978.
- . *Understanding Copyright and Related Rights*. Second edition., 2016. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.
- World Intellectual Property Organization (WIPO). “Berne Convention.” Accessed June 7, 2024. <https://www.wipo.int/wipolex/en/text/283698>.
- World Intellectual Property Organization (WIPO). “WIPO Performances and Phonograms Treaty.” Accessed June 7, 2024. <https://www.wipo.int/wipolex/en/text/295578>.
- World Trade Organization. “WTO | Intellectual Property (TRIPS) - Agreement Text - Contents.” Accessed June 7, 2024. https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm.

“WTO | Intellectual Property (TRIPS) - Agreement Text - Standards.” Accessed June 20, 2024. https://www.wto.org/english/docs_e/legal_e/27-trips_04_e.htm#.

Yetim, Servet. “Yargı Kararlarına Erişim Hakkı.” *Uyuşmazlık Mahkemesi Dergisi* 4, no. 4 (April 1, 2014). <https://doi.org/10.18771/umd.90664>.

Yılmaztekin, Hasan Kadir. *Yapay Zekanın Eser Sahipliği*. Ankara: Adalet Yayınları, 2021.

Your Europe. “Database Protection in the EU.” Accessed December 29, 2023. https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_en.htm.