

Nr. 2 Kneschke/LAION

Landgericht Hamburg 27 september 2024

Az: 310 O 227/23

Eerste vonnis in Europa over de TDM-beperkingen in art. 3 en 4 DSM-richtlijn. Het downloaden door LAION van een op een website aangetroffen foto voor het samenstellen van een dataset die gebruikt kan worden voor AI-trainingsdoeleinden is op grond van (de Duitse implementatie van) art. 3 DSM-R1 toegestaan, omdat LAION als onderzoeksinstelling zonder winstoogmerk kwalificeert en het downloaden ertoe dient om de gegevens in de dataset te verifiëren. Daarbij is aan de driestappentoets voldaan. In obiter dictum overweegt het Landgericht dat een voorbehoud als bedoeld in art. 4 lid 3 DSM-R1 door een licentiehouder rechtsgeldig kan worden gemaakt, en dat – afhankelijk van de stand van de techniek – een in natuurlijke taal gestelde gebruiksbeperking op een site als een “machine-leesbaar” voorbehoud zou kunnen gelden.

Tatbestand

- Der Beklagte ist ein Verein, der mit Gründungsversammlung vom 07.07.2021 gegründet wurde (Protokoll in Anlage B7, Satzung in Anlage B1). Die konkrete Zweckrichtung der Tätigkeit des Beklagten steht zwischen den Parteien in Streit.
- Der Beklagte stellt unter der Bezeichnung “L.” ein sogenanntes Dataset für Bild-TextPaare öffentlich kostenfrei zur Verfügung. Es handelt sich dabei um eine Art Tabellen-dokument, das Hyperlinks zu im Internet öffentlich abrufbaren Bildern bzw. Bilddateien sowie weitere Informationen zu den entsprechenden Bildern enthält, darunter eine Bildbeschreibung (auch Alternativtext genannt), die Auskunft über den Inhalt des Bildes in Textform gibt. Der Datensatz umfasst 5,85 Milliarden entsprechende Bild-Text-Paare. Der Datensatz kann für das Trainieren sog. generativer Künstlicher Intelligenz genutzt werden.
- Die Erstellung des Datensatzes erfolgte nach Gründung des Beklagten in der zweiten Jahreshälfte 2021. Dazu hatte der Beklagte auf einen bereits vorhandenen Datensatz der C. C. F. aus den USA (www.c..org) zurückgegriffen, der für eine Art zufälligen Querschnitt der im Internet auffindbaren Bilder die jeweiligen URLs nebst textlicher Beschreibung des jeweiligen Bildinhalts enthielt. Der Beklagte extrahierte sodann die URLs zu den Bildern aus diesem Datensatz und lud die Bilder von ihrem jeweiligen Speicherort herunter. Im Anschluss wurden die Bilder beim Beklagten durch eine Software darauf geprüft, ob die im vorbestehenden Datensatz bereits vorhandene Beschreibung des Bildinhalts tatsächlich mit dem auf dem Bild zu sehenden Inhalt übereinstimmte. Bilder, bei denen Text

und Bildinhalt nicht hinreichend übereinstimmten, wurden herausgefiltert. Für die verbleibenden Bilder wurden die Meta-Daten, insbesondere die URL des Speicherorts des Bildes und die Bildbeschreibung, extrahiert und in einen neu geschaffenen Datensatz, den L., aufgenommen. Ob im Anschluss die heruntergeladenen Bilddateien wieder gelöscht wurden, steht zwischen den Parteien – jedenfalls in Bezug auf die streitgegenständliche Fotografie – in Streit.

- Im Rahmen des vorgenannten Prozesses wurde auch das streitgegenständliche Bild erfasst, heruntergeladen, analysiert und mit seinen Meta-Daten in den Datensatz L. aufgenommen. Konkret heruntergeladen wurde dabei eine auf der Webseite der Bildagentur B. (<https://www.b..com>) eingestellte, mit einem Wasserzeichen der Fotoagentur B. versehene Bilddatei.
- Auf der Webseite der Bildagentur B. befand sich mindestens seit dem 13.01.2021 auf der Unterseite <https://www.b..com/de/usage.html> folgender Text:
- „RESTRICTIONS
- YOU MAY NOT:
- [...]
18. Use automated programs, applets, bots or the like to access the B..com website or any content thereon for any purpose, including, by way of example only, downloading Content, indexing, scraping or caching any content on the website.“

[...]

Entscheidungsgründe

- [...]
- [...]
- Die erfolgte Vervielfältigung ist nicht durch die Schrankenregelung des § 44a UrhG gedeckt.
- Danach sind vorübergehende Vervielfältigungshandlungen zulässig, die flüchtig oder begleitend sind und einen integralen und wesentlichen Teil eines technischen Verfahrens darstellen und deren alleiniger Zweck es ist, eine Übertragung in einem Netz zwischen Dritten durch einen Vermittler oder eine rechtmäßige Nutzung eines Werkes

oder sonstigen Schutzgegenstands zu ermöglichen, und die keine eigenständige wirtschaftliche Bedeutung haben.

60. Die vorliegend erfolgte Vervielfältigung war bereits weder flüchtig noch begleitend.

a)

61. Flüchtig i.S.d. § 44a UrhG ist eine Vervielfältigung dann, wenn ihre Lebensdauer auf das für das ordnungsgemäße Funktionieren des betreffenden technischen Verfahrens Erforderliche beschränkt ist, wobei dieses Verfahren derart automatisiert sein muss, dass es diese Handlung automatisch, d.h. ohne Beteiligung einer natürlichen Person löscht, sobald ihre Funktion, die Durchführung eines solchen Verfahrens zu ermöglichen, erfüllt ist (EuGH, Urt. v. 16.07.2009, Az. C-5/08 – Infopaq/Danske Dagblades Forening, Rn. 64 (juris) zu Art. 5 Abs. 1 DSM-RL).

62. Soweit sich der Beklagte insoweit darauf beruft, dass im Rahmen des von ihm durchgeführten Analyseverfahrens die Dateien „automatisch“ gelöscht worden seien, vermag dies eine Flüchtigkeit der Vervielfältigung im vorgenannten Sinne nicht zu begründen. Abgesehen davon, dass der Beklagte nichts zu der konkreten Dauer der Speicherung vorgetragen hat, erfolgte die Löschung gerade nicht „nutzunabhängig“, sondern aufgrund einer entsprechenden bewussten Programmierung des Analyseprozesses durch den Beklagten.

b)

63. Begleitend i.S.d. § 44a UrhG ist eine Vervielfältigung dann, wenn sie gegenüber dem technischen Verfahren, dessen Teil sie ist, weder eigenständig ist noch einem eigenständigen Zweck dient (EuGH, Urt. v. 05.06.2014, Az. C-360/13, Rn. 43 (juris)).

64. Im vorliegenden Fall erfolgte ein gezieltes Herunterladen der Bilddateien, um sie mittels einer spezifischen Software zu analysieren. Damit ist das Herunterladen kein bloß begleitender Prozess zu der durchgeführten Analyse, sondern ein der Analyse vorgelagerter bewusster und aktiv gesteuerter Beschaffungsprozess.

2.

65. Ob sich der Beklagte auf die Schrankenregelung des § 44b UrhG berufen kann, erscheint im vorliegenden Fall durchaus als zweifelhaft. Zwar unterfällt der von dem Beklagten vorgenommene Download grundsätzlich der Schrankenregelung des § 44b Abs. 2 UrhG, insbesondere erfolgte er zum Zwecke des Text und Data Mining im Sinne des § 44b Abs. 1 UrhG (im Folgenden lit. a). Allerdings spricht – ohne dass dies vorliegend einer abschließenden Entscheidung bedürfte – Einiges dafür, dass aufgrund eines wirksam erklärten Nutzungsvorbehalt im Sinne des § 44b Abs. 3 die Vervielfältigungshandlung nicht bereits nach § 44b Abs. 2 UrhG zulässig war (im Folgenden lit. b).

a)

66. Die streitgegenständliche Vervielfältigungshandlung unterfällt grundsätzlich der Schrankenregelung des § 44b Abs. 2 UrhG.

67. (1) Der streitgegenständliche Download erfolgte zum Zwecke des Text und Data Mining im Sinne des § 44b Abs. 1 UrhG. Danach ist Text und Data Mining die automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen. Jedenfalls für die vorliegend streitgegenständliche Vervielfältigungshandlung ist dies zu bejahen (nachfolgend (a)); eine teleologische Reduktion des Schrankentatbestands kommt insofern nicht in Betracht (unten (b)).

68. Vorliegend keiner Entscheidung bedarf daher die weitere, im Schriftum eingehend diskutierte Frage, ob das Training von Künstlicher Intelligenz in seiner Gesamtheit der Schrankenregelung des § 44b UrhG unterfällt oder nicht (eingehend zum Meinungsstand BeckOK UrhR/Bomhard, 42. Ed. 15.2.2024, UrhG § 44b Rn. 11a-11b m.w.N.; dazu auch eingehend die als Anlage K11 vorgelegte, im Auftrag der Initiative Urheberrecht erstellte Studie „Urheberrecht & Training generativer KI-technologische und rechtliche Grundlagen“).

69. (a) Der Beklagte hat die Vervielfältigungshandlung zum Zwecke der Gewinnung von Informationen über „Korrelationen“ im Wortsinn des § 44b Abs. 1 UrhG vorgenommen. Der Beklagte hat das streitgegenständliche Lichtbild von seinem ursprünglichen Speicherort heruntergeladen, um mittels einer bereits verfügbaren Software – offenbar der Anwendung CLIP von OpenAI – den Bildinhalt mit der zu dem Text bereits hinterlegten Bildbeschreibung abzuleichen. Diese Analyse der Bilddatei zum Abgleich mit einer vorbestehenden Bildbeschreibung stellt ohne Weiteres eine Analyse zum Zwecke der Gewinnung von Informationen über „Korrelationen“ (nämlich der Frage der Nicht-/Übereinstimmung von Bildern und Bildbeschreibungen) dar. Dass der Beklagte die in den Datensatz L. aufgenommenen Bilder auf diese Art und Weise analysiert hat, wurde klägerseits als solches nicht bestritten.

[...]

71. (b) Die streitgegenständliche Vervielfältigungshandlung ist auch nicht im Wege der teleologischen Reduktion der Schrankenregelung des § 44b UrhG aus dieser auszuschließen.

72. Soweit eine Herausnahme der Vervielfältigung von Daten zum Zwecke des KI-Trainings im Wege der teleologischen Reduktion im Schriftum vereinzelt mit der Begründung befürwortet wird, dass § 44b UrhG nur die Erschließung „in den Daten verborgener Informationen“, nicht aber die Nutzung „des Inhalts der geistigen Schöpfung“ erfasse (Schack, NJW 2024, 113; in diese Richtung auch Dor-

mis/Stober, Urheberrecht und Training generativer KI-Modelle, Anlage K11, S. 67 ff. mit einer Differenzierung zwischen Semantik und Syntax), bestehen Zweifel, ob dies zu überzeugen vermag; denn dabei wird nicht ausreichend deutlich, worin bei digitalisierten Werken der Unterschied zwischen "in den Daten verborgenen Informationen" und "dem Inhalt der geistigen Schöpfung" liegen soll.

[...]

79. Davon abgesehen hat jedenfalls der aktuelle europäische Gesetzgeber der KI-Verordnung (Verordnung (EU) 2024/1689 vom 13.06.2024, ABl. L v. 12.07.2024 S. 1) unzweifelhaft zum Ausdruck gebracht, dass auch die Erstellung von zum Training von künstlichen neuronalen Netzen bestimmten Datensätzen der Schrankenregelung des Art. 4 DSM-RL unterfällt. Denn nach Art. 53 Abs. 1 lit. c KI-Verordnung sind Anbieter von KI-Modellen mit allgemeinem Verwendungszweck verpflichtet, eine Strategie insbesondere zur Ermittlung und Einhaltung eines gemäß Art. 4 Abs. 3 DSM-RL geltend gemachten Rechtsvorbehalts vorzusehen.

80. Dass auch die Erstellung von zum Training von künstlichen neuronalen Netzen bestimmten Datensätzen der Schrankenregelung des Art. 4 DSM-RL unterfalle, entspricht im Übrigen auch der Einschätzung des deutschen Gesetzgebers im Rahmen der Umsetzung der vorgenannten Schrankenbestimmung im Jahr 2021 (Begr. RegE BT-Drs. 19/27426, S. 60).

81. (c) Auch der in Art. 5 Abs. 5 InfoSoc-RL (i.V.m. Art. 7 Abs. 2 S. 1 DSM-RL) verankerte sog. 3-Stufen-Test rechtfertigt letztlich keine andere Beurteilung. Danach dürfen die normierten Ausnahmen nur in bestimmten Sonderfällen angewandt werden, in denen die normale Verwertung des Werks oder des sonstigen Schutzgegenstands nicht beeinträchtigt wird und die berechtigten Interessen des Rechtsinhabers nicht ungebührlich verletzt werden. Diese Anforderungen sind im vorliegenden Fall erfüllt.

82. Die vorliegend urheberrechtlich relevante Vervielfältigung ist auf den Zweck der Analyse der Bilddateien auf ihre Übereinstimmung mit einer vorbestehenden Bildbeschreibung nebst anschließender Einstellung in einen Datensatz beschränkt. Dass durch diese Nutzung die Verwertungsmöglichkeiten der jeweils betroffenen Werke beeinträchtigt werden würden, ist nicht ersichtlich und wird auch klägerseits nicht behauptet.

83. Zwar mag der auf diese Weise erstellte Datensatz nachfolgend zum Trainieren künstlicher neuronaler Netze genutzt werden können und die dabei entstehenden KI-generierten Inhalte mögen in Konkurrenz zu den Werken (menschlicher) Urheber treten können. Das allein rechtfertigt es jedoch noch nicht, bereits in der Erstellung der Trainingsdatensätze eine Beeinträchtigung auch der Verwertungsrechte an Werken i.S.v. Art. 5 Abs.5 InfoSoc-RL zu erblicken. Dies hat schon allein deshalb zu gelten, weil die

Berücksichtigung bloß zukünftiger, derzeit noch gar nicht im Einzelnen absehbarer technischer Entwicklungen keine rechtssichere Abgrenzung zulässiger von unzulässigen Nutzungen erlaubt (vgl. ähnlich oben (b)).

84. Da eine Verwendung von im Wege des Text und Data Mining gewonnenen Erkenntnissen zum Trainieren künstlicher neuronaler Netze, die dann in Konkurrenz zu Urhebern treten können, auf Grundlage der aktuellen technologischen Entwicklung im Zweifel nie ausgeschlossen werden kann, würde die Gegenauffassung in letzter Konsequenz sogar dazu nötigen, das Text und Data Mining i.S.d. § 44b UrhG letztlich in seiner Gänze zu untersagen; eine solche vollständige Außerkraftsetzung der Schrankenregelung liefe aber der gesetzgeberischen Intention offenkundig zuwider und kann daher kein tragbares Auslegungsergebnis darstellen.

85. (2) Die von dem Beklagten heruntergeladene Bilddatei war auch – was der Kläger im Übrigen auch nicht in Abrede stellt – rechtmäßig zugänglich i.S.d. § 44b Abs. 2 S. 1 UrhG.

86. „Rechtmäßig zugänglich“ in diesem Sinne ist ein Werk insbesondere dann, wenn es frei im Internet zugänglich ist (Begr. RegE BT-Drucks. 19/27426, S. 88).

87. Hiervon ist für das von dem Beklagten heruntergeladene Bild auszugehen. Anders als klägerseits zunächst vorgetragen, hat der Beklagte nicht das in dem in der Klageschrift zunächst formulierten Unterlassungsantrag wiedergegebene „Originalbild“ – das von der Bildagentur B. nur bei Erwerb einer Lizenz zur Verfügung gestellt worden wäre –, sondern eine mit einem Wasserzeichen der Bildagentur versehene Fassung des Bildes heruntergeladen. Hierbei handelte es sich ersichtlich um das auf der Agenturseite quasi zu Werbezwecken eingestellte Vorschaubild. Dieses mit dem Wasserzeichen versehene Vorschaubild wurde von der Agentur aber gerade frei zugänglich ins Internet gestellt.

b)

88. Allerdings spricht einiges dafür, dass vorliegend die Schrankenregelung des § 44b Abs. 2 UrhG – ohne dass dies einer abschließenden Entscheidung bedürfte – nicht eingreift, da ein wirksam erklärter Nutzungsvorbehalt im Sinne von Abs. 3 der Vorschrift vorlag; insbesondere dürfte der auf der Webseite B..com unstreitig erklärte Nutzungsvorbehalt wohl den Anforderungen an eine Maschinenlesbarkeit i.S.d. § 44b Abs. 3 S. 2 UrhG genügen.

89. (1) Es spricht manches dafür, dass der auf der Webseite der Agentur ausgesprochene Nutzungsvorbehalt durch eine hierzu berechtigte Person ausgesprochen wurde und der Kläger sich hierauf auch zum Schutz seiner eigenen Rechte berufen kann.

90. Nach dem Wortlaut des § 44b Abs. 3 UrhG kann "der Rechtsinhaber" den Nutzungsvorbehalt aussprechen. Zu

beachten sind also nicht allein Vorbehaltserklärungen des Urhebers selbst, sondern auch nachfolgender Rechteinhaber, seien sie Rechtsnachfolger oder Inhaber von vom Urheber abgeleiteten Rechten. Nach dem ohne Weiteres schlüssigen Vortrag des Klägers (Protokoll v. 11.07.2024 S. 3, Bl. 122 d.A.) hatte er der Bildagentur B. weiterlizenzierbare einfache Nutzungsrechte an dem Originalbild eingeräumt. Die Bildagentur war danach selbst Rechteinhaberin an den auf ihrer Seite eingestellten Bildern und konnte daher ohne Weiteres einen Nutzungsvorbehalt nach § 44b Abs. 3 UrhG aussprechen; dass insofern dinglich wirkende Vereinbarungen im Vertragsverhältnis zwischen dem Kläger und der Bildagentur entgegengestanden hätten, ist weder ersichtlich noch geltend gemacht worden.

91. Der Kläger ist wohl auch berechtigt, sich auf diese Vorbehaltserklärung seiner Lizenznehmerin zu berufen. Wirtschaftlich betrachtet fand hier die Auswertung des streitgegenständlichen Originalfotos über die Agentur statt. Damit lag in der Praxis die konkrete Entscheidung, welcher Dritte die Berechtigung zu welcher Nutzung erhalten sollte, bei der Agentur; Abschlusszwang bestand für diese nicht. In einer solchen Situation spricht aus Sicht der Kammer vieles dafür, dass sich der Urheber bei der Geltendmachung bei ihm verbliebener Verbietungsrechte auf einen von seinem Lizenznehmer erklärten Vorbehalt nach § 44b Abs. 3 UrhG berufen darf.

92. (2) Auch der Einwand des Beklagten, dass das in den AGB der Agentur gegenüber deren Kunden ausgesprochene Nutzungsverbot für Webcrawlers schon in zeitlicher Hinsicht nicht "in Bezug auf § 44b Abs. 3 UrhG" formuliert sein könne, ist unerheblich. Für die Rechtswirkungen der Erklärung ist es keine Voraussetzung, dass sie bewusst mit Blick auf eine bestimmte Gesetzesfassung erklärt wird.

93. (3) Der Vorbehalt ist auch hinreichend klar formuliert. Art. 4 Abs. 3 DSM-RL verlangt eine ausdrückliche Erklärung des Nutzungsvorbehalts. Dieses Ausdrücklichkeitserfordernis ist mithin bei richtlinienkonformer Auslegung des § 44b Abs. 3 UrhG mitzuberücksichtigen (so auch Begr. RegE BT-Drs. 19/27426, 89). Der erklärte Vorbehalt muss damit sowohl expressis verbis (nicht konkludent) als auch so zielgenau (konkret-individuell) erklärt werden, dass er zweifelsfrei einen bestimmten Inhalt und eine bestimmte Nutzung erfasst (Hamann, ZGE 16 (2024), S. 134). Diese Anforderungen genügt der auf der Webseite der Bildagentur B. formulierte Nutzungsvorbehalt ohne Weiteres.

94. Soweit darüber hinaus argumentiert wird, dass ein für sämtliche auf einer Webseite eingestellte Werke erklärter Nutzungsvorbehalt dem Ausdrücklichkeitserfordernis des § 44b Abs. 3 UrhG widerspreche (so in Erweiterung der eigenen abstrakten Herleitung Hamann, a.a.O., S. 148), vermag dies nicht zu überzeugen. Denn auch der explizit für sämtliche auf einer Webseite eingestellte Werke erklärte Vorbehalt ist in seiner Reichweite und seinem Inhalt nach zweifelsfrei bestimmbar und damit ausdrücklich erklärt.

95. (4) Schließlich dürfte auch Einiges dafür sprechen, dass der Nutzungsvorbehalt den Anforderungen an eine Maschinenlesbarkeit i.S.d. § 40b Abs. 3 S. 2 UrhG genügt.

96. Dabei wird man zwar den Begriff der Maschinenlesbarkeit im Hinblick auf den ihm zugrunde liegenden gesetzgeberischen Willen, eine automatisierte Abfrage durch Webcrawler zu ermöglichen (vgl. Begr. RegE BT-Drucks. 19/27426, S. 89), durchaus im Sinne einer "Maschinenverständlichkeit" auszulegen haben (eingehend zum Meinungsstand Hamann, a.a.O., S. 113, 128 ff.).

97. Die Kammer neigt allerdings dazu, als "maschinenverständlich" auch einen allein in "natürlicher Sprache" verfassten Nutzungsvorbehalt anzusehen (anders als die wohl überwiegende Auffassung im Schrifttum, s. Hamann, a.a.O., S. 131 ff., 146 ff. m.w.N. zum Meinungsstand, wobei dort auf einen Beitrag der hiesigen Beklagtenvertreter, nämlich auf Akinci/Heidrich, IPRB 2023, 270, 272 verwiesen wird, die offenbar ebenfalls die Auffassung der Kammer vertreten; der Beitrag war der Kammer allerdings bis zur Urteilsabfassung nicht unmittelbar zugänglich). Allerdings wird man die Frage, ob und unter welchen konkreten Voraussetzungen ein in "natürlicher Sprache" erklärter Vorbehalt auch als "maschinenverständlich" angesehen werden kann, stets in Abhängigkeit von der zum jeweils relevanten Werknutzungszeitpunkt bestehenden technischen Entwicklung beantworten müssen.

98. Dementsprechend hat auch der europäische Gesetzgeber im Rahmen der KI-Verordnung festgelegt, dass Anbieter von KI-Modellen eine Strategie insbesondere zur Ermittlung und Einhaltung eines gemäß Art. 4 Abs. 3 DSM-RL geltend gemachten Rechtvorbehalts "auch durch modernste Technologien" vorzuhalten haben (Art. 53 Abs. 1 lit. c KI-Verordnung). Zu diesen "modernsten Technologien" gehören aber unzweideutig gerade auch KI-Anwendungen, die in der Lage sind, in natürlicher Sprache geschriebenen Text inhaltlich zu erfassen (so offenbar insbesondere auch die hiesigen Beklagtenvertreter Akinci/Heidrich in dem der Kammer nicht unmittelbar zugänglichen Beitrag IPRB 2023, 270, 272, hier zit. nach Hamann, a.a.O. S. 148, dieser im Übrigen diese Möglichkeit in technischer Hinsicht grds. bejahend, a.a.O.). Es spricht insoweit alles dafür, dass der Gesetzgeber der KI-Verordnung mit seinem Hinweis auf "modernsten Technologien" gerade solche KI-Anwendungen im Blick hatte.

[...]

100. Ferner wird gegen das von der Kammer erwogene, weitere Verständnis des Begriffes der "Maschinenlesbarkeit" auch eingewandt, dieser Begriff werde vom europäischen Gesetzgeber in anderem Zusammenhang enger verstanden. Verwiesen wird in diesem Zusammenhang auf den Erwägungsgrund 35 der PSI-Richtlinie (RL (EU) 2019/1024), der für eine "Maschinenlesbarkeit" im Sinne dieser Richtlinie u.a. eine "einfache" Erkennbarkeit verlange (so BeckOK UrhR/Bomhard, 42. Ed. 15.2.2024, UrhG §

44b Rn. 31 m.w.N.); dies könnte für einen nur in natürlicher Sprache formulierten Vorbehalt nicht angenommen werden. Eine solche Argumentation setzt allerdings voraus, dass die Begrifflichkeiten beider Richtlinien in gleicher Weise verstanden werden müssen. Die Kammer hat Zweifel, ob eine solche Gleichsetzung der Begrifflichkeiten überzeugen kann, denn die Richtlinien haben unterschiedliche Zielrichtungen: Während die PSI-Richtlinie den rein einseitigen Zugang der Öffentlichkeit zu Informationen bzw. die rein einseitige Verpflichtung der öffentlichen Hand zur Veröffentlichung bestimmter Informationen zum Gegenstand hat, geht es im Rahmen von Art. 4 Abs. 3 DSM-RL um einen Ausgleich zwischen den Interessen der Nutzer des Text und Data Mining (dieses möglichst einfach und möglichst rechtssicher betreiben zu können) und den Interessen der Rechteinhaber (ihre Rechte möglichst einfach und möglichst effektiv zu sichern). Dieser Interessenausgleich kann nach Auffassung der Kammer nicht einseitig zugunsten der Nutzer des Text und Data Mining gelöst werden, indem allein die für diese denkbar einfachste technische Lösung als ausreichend für die Wirksamkeit eines ausgesprochenen Nutzungsvorbehalts erachtet wird. Gegen ein solches Verständnis spräche auch die Wertung des Gesetzgebers der DSMRL, der im Erwägungsgrund 18 gerade nicht die Erklärung eines Vorbehalts "in möglichst einfach auszulesender Weise" fordert, sondern lediglich "in angemessener Weise". Und auch der deutsche Umsetzungsgesetzgeber verlangt lediglich eine Erklärung in einer Weise, die "den automatisierten Abläufen beim Text und Data Mining angemessen" ist (Begr. RegE BT-Drucks. 19/27426, S. 89).

101. Es wäre zudem aus Sicht der Kammer ein gewisser Wertungswiderspruch, den Anbietern von KI-Modellen einerseits über die Schranke in § 44b Abs. 2 UrhG die Entwicklung immer leistungsfähigerer textverstehender und -kreierender KI-Modelle zu ermöglichen, ihnen aber andererseits im Rahmen der Schranken-Schranke von § 44b Abs. 3 S. 2 UrhG die Anwendung bereits bestehender KI-Modelle nicht abzuverlangen.

[...]

3.

103. Der Beklagte kann sich hinsichtlich der streitgegenständlichen Vervielfältigung jedoch auf die Schrankenregelung des § 60d UrhG berufen.

104. Danach sind Vervielfältigungen für Text und Data Mining für Zwecke der wissenschaftlichen Forschung durch Forschungsorganisationen zulässig.

a)

105. Die Vervielfältigung erfolgte – wie dargelegt – für den Zweck des Text und Data Mining i.S.d. § 44b Abs. 1 UrhG. Sie erfolgte darüber hinaus auch zu Zwecken der wissenschaftlichen Forschung i.S.d. § 60d Abs. 1 UrhG.

106. Wissenschaftliche Forschung bezeichnet allgemein das methodisch-systematische Streben nach neuen Erkenntnissen (Spindler/Schuster/Anton, 4. Aufl. 2019, UrhG § 60c Rn.3; BeckOK UrhR/Grübler, 42. Ed. 1.5.2024, UrhG § 60c Rn. 5; Dreier/Schulze/Dreier, 7. Aufl. 2022, UrhG § 60c Rn. 1). Der Begriff der wissenschaftlichen Forschung ist, indem er bereits das methodisch-systematische "Streben" nach neuen Erkenntnissen ausreichen lässt, nicht so eng zu verstehen, dass er nur die unmittelbar mit der Gewinnung von Erkenntnisgewinn verbundenen Arbeitsschritte erfassen würde; vielmehr genügt es, dass der in Rede stehende Arbeitsschritt auf einen (späteren) Erkenntnisgewinn gerichtet ist, wie es z.B. bei zahlreichen Datensammlungen der Fall ist, die zunächst durchgeführt werden müssen, um anschließend empirische Schlussfolgerungen zu ziehen. Insbesondere setzt der Begriff der wissenschaftlichen Forschung auch keinen späteren Forschungserfolg voraus.

107. Danach kann – entgegen der Auffassung des Klägers – auch bereits die Erstellung eines Datensatzes der streitgegenständlichen Art, der Grundlage für das Trainieren von KI-Systemen sein kann, durchaus als wissenschaftliche Forschung im vorstehenden Sinne anzusehen sein. Zwar mag die Erstellung des Datensatzes als solche noch nicht mit einem Erkenntnisgewinn verbunden sein; sie ist aber grundlegender Arbeitsschritt mit dem Ziel, den Datensatz zum Zwecke späteren Erkenntnisgewinns einzusetzen. Dass eine solche Zielsetzung auch im vorliegenden Fall bestand, kann bejaht werden. Dafür genügt es, dass der Datensatz – unstreitig – kostenfrei veröffentlicht und damit gerade (auch) auf dem Gebiet künstlicher neuronaler Netze Forschenden zur Verfügung gestellt wurde. Ob der Datensatz – wie es der Kläger hinsichtlich der Dienste DALL-E 2 und MidJourney behauptet – daneben auch von kommerziellen Unternehmen zum Training bzw. zur Weiterentwicklung ihrer KI-Systeme genutzt wird, ist schon deshalb unerheblich, weil auch die Forschung kommerzieller Unternehmen noch Forschung – wenn auch nicht als solche nach §§ 60c f. UrhG privilegiert – ist.

108. Auf die zwischen den Parteien streitige Frage, ob der Beklagte über die Erstellung entsprechender Datensätze hinaus auch wissenschaftliche Forschung in Gestalt der Entwicklung eigener KI-Modelle tätigt, kommt es vor diesem Hintergrund nicht an.

b)

109. Der Beklagte verfolgt auch nicht kommerzielle Zwecke i.S.d. § 60d Abs. 2 Nr. 1 UrhG.

110. Für die Frage, ob Forschung nicht kommerziell ist, kommt es allein auf die konkrete Art der wissenschaftlichen Tätigkeit an, während Organisation und Finanzierung der Einrichtung, in der die Forschung erfolgt, unbedachtlich sind (ErwGr 42 InfoSoc-RI).

111. Die nicht-kommerzielle Zweckverfolgung des Beklagten in Bezug auf die streitgegenständliche Erstellung des

Datensatzes L. ergibt sich dabei bereits daraus, dass der Beklagte diesen unstreitig kostenfrei öffentlich zur Verfügung stellt. Dass die Entwicklung des streitgegenständlichen Datensatzes darüber hinaus zumindest auch der Entwicklung eines eigenen kommerziellen Angebots des Beklagten dienen würde (vgl. zu diesem Kriterium BeckOK IT-Recht/Paul, 14. Ed. 1.4.2024, UrhG § 60d Rn. 10), ist weder klägerseits vorgetragen noch sonst ersichtlich. Dass der streitgegenständliche Datensatz auch von kommerziell tätigen Unternehmen zum Training bzw. zur Weiterent-

wicklung ihrer KI-Systeme genutzt werden mag, ist für die Einordnung der Tätigkeit des Beklagten hingegen ohne Relevanz. Allein der Umstand, dass einzelne Mitglieder des Beklagten neben ihrer Tätigkeit für den Verein auch bezahlten Tätigkeiten bei solchen Unternehmen nachgehen, genügt nicht, die Tätigkeit dieser Unternehmen dem Beklagten als eigene zuzurechnen.

[...]

Noot

Bernt Hugenholtz

Prof. mr. P.B. Hugenholtz is hoogleraar emeritus aan de Universiteit van Amsterdam, Instituut voor Informatierecht, en redacteur van dit blad.

Dit vonnis van het Hamburgse Landgericht (de Duitse civiele rechter in eerste instantie) is de eerste rechterlijke uitspraak in Europa over de uitleg van de tekst- en datamining ('TDM') bepalingen in de DSM-richtlijn. De uitspraak heeft daarom – vooral met het oog op haar implicaties voor het trainen van AI-modellen – veel aandacht gekregen in de internationale vakpers.

Eiser in deze zaak, beroepsfotograaf Robert Kneschke, klaagt dat LAION, een non-profit organisatie die 'open' datasets om niet beschikbaar stelt voor AI-onderzoek en -toepassingen,¹ auteursrechtinbreuk heeft gepleegd door een van zijn foto's van de website van zijn fotoagentschap te downloaden. De dataset van LAION bouwt voort op een oudere dataset afkomstig van Common Crawl,² bestaande uit hyperlinks naar beeldbestanden met summiere beschrijvingen. LAION heeft op basis hiervan een nieuwe dataset gecreëerd waarbij de beeldbestanden tijdelijk zijn gedownload om de beschrijvingen te verifiëren. Het LAION-bestand bestaat uit een combinatie van url's en beeldbeschrijvingen, maar bevat geen beeldbestanden. LAION ontwikkelt zelf geen AI-modellen, zoals de grote spelers op dit terrein, OpenAI, Google en Meta dat wel doen. De Hamburgse zaak is dus niet de grote auteursrecht contra AI-zaak, waarnaar in Europa al geruime tijd wordt uitgezien.³

De Hamburgse rechtsbank wijst de klacht van Kneschke af. Volgens het Landgericht beroeft LAION zich met succes op art. 60a van de Duitse auteurswet (het equivalent van art. 15n Aw), de bepaling waarin art. 3 van de DSM-richtlijn is omgezet. Op grond hiervan is het "onderzoeksorganisaties en instellingen voor cultureel erfgoed" toegestaan "om met het oog op wetenschappelijk onderzoek tekst- en datamining te verrichten" (art. 3 lid 1 DSM-RI). Daarbij hanteert de Hamburgse

rechtsbank een ruime opvatting van het begrip 'wetenschappelijk onderzoek'. Volgens de rechter volstaat dat de activiteit in kwestie uiteindelijk gericht is op kennisvermeerdering (r.o. 106). De dataset is immers bedoeld om gratis aan AI-onderzoekers beschikbaar te worden gesteld. Dat de dataset ook door commerciële AI-ontwikkelaars voor trainingsdoeleinden kan worden benut, maakt geen verschil; immers, commerciële research is ook een vorm van onderzoek, aldus het Landgericht (r.o. 107).

Interessanter dan deze overwegingen is het uitvoerige *obiter dictum* over art. 44b van de Duitse wet (in Nederland art. 150 Aw), de transpositie van art. 4 DSM-richtlijn. Op grond van deze, voor de AI-praktijk veel belangrijkere bepaling is tekst- en datamining in algemene zin toegestaan ongeacht het nagestreefde doel, tenzij de rechthebbende uitdrukkelijk een voorbehoud ('opt-out') heeft gemaakt. Hoewel de Duitse rechter met zijn oordeel over art. 3 DSM-RI had kunnen volstaan, gaat het Landgericht uitvoerig in op enkele vragen van uitleg van art. 4 DSM die in de literatuur en praktijk zijn opgeworpen.

Het Landgericht stelt voorop dat de litigieuze activiteit – het downloaden van de foto van Kneschke en het automatisch vergelijken van het beeld met de beschrijving – te beschouwen is als een vorm van 'tekst- en data mining' gedefinieerd in art. 2 (2) DSM-RI: "een geautomatiseerde analysetechniek die gericht is op de ontleding van tekst en gegevens in digitale vorm om informatie te genereren zoals, maar niet uitsluitend, patronen, trends en onderlinge verbanden" (r.o. 69).

Vervolgens bespreekt de Hamburgse rechter de door enkele Duitse schrijvers verdedigde opvatting dat de TDM-

¹ <https://laion.ai/about/>.

² <https://commoncrawl.org/>.

³ Inmiddels heeft de Duitse CBO GEMA bekend gemaakt een inbreukprocedure tegen OpenAI te beginnen; zie <https://www.gema.de/en/w/gema-files-lawsuit-against-openai>.

beperkingen eng zijn uit te leggen in die zin dat zij enkel ten doel hebben de in de ‘geminide’ werken verborgen informatie te ontsluiten - en niet “de inhoud van de geestelijke schepping” (dat wil zeggen: de expressie of vorm van het werk). In deze opvatting zou het ‘minen’ voor AI-trainingsdoeleinden buiten de reikwijdte van de beperking vallen; immers, het AI-model onttrekt niet enkel informatie aan de gebruikte werken, maar ‘leert’ ook van de vorm ervan.⁴ De rechter is hiervan echter niet overtuigd, omdat het onderscheid tussen informatie en expressie moeilijk te maken is (r.o. 72). Wat hiervan verder ook zij, de AI Act (Verordening (EU) 2024/1689) die op 13 juni 2024 is aangenomen, heeft aan iedere twijfel op dit punt een einde gemaakt door te bepalen dat ontwikkelaars van algemene AI-modellen zich aan art. 4 DSM-RI hebben te houden, aldus het Landgericht (r.o. 79). Ook de Duitse wetgever is er bij de omzetting van de DSM-richtlijn van uitgegaan dat art. 4 DSM-RI van toepassing is op het gebruik van werken voor trainingsdoeleinden (r.o. 80).

Het Landgericht wijdt vervolgens enkele overwegingen aan de ‘driestappentoets’ van art. 5 lid 5 Infosoc-richtlijn, waaraan ook de TDM-beperkingen uit de DSM-richtlijn moeten voldoen (art. 7 lid 2 DSM-RI). Volgens de rechtsbank valt niet in te zien hoe de totstandkoming van de LAION-dataset de normale exploitatie van de geminide werken schaadt (r.o. 82). “Weliswaar kan de op deze manier gecreerde dataset vervolgens worden gebruikt om kunstmatige neurale netwerken te trainen, en kan de resulterende AI-gegenereerde inhoud mogelijk concurreren met het werk van (menselijke) auteurs. Dit alleen is echter geen reden om het creëren van de trainingsdatasets te beschouwen als een aantasting van de exploitatierechten van werken in de zin van artikel 5 lid 5 van de InfoSoc-richtlijn” (r.o. 83). Als met ieder – voor de auteurs mogelijk schadelijk – toekomstig gebruik van door TDM gegenereerde datasets rekening zou moeten worden gehouden, blijft er van de TDM-beperking niets over, aldus de Hamburgse rechter (r.o. 84).

Vervolgens toetst het Landgericht de activiteiten van LAION aan de vereisten van art. 44b van de Duitse wet (art. 4 DSM-RI). De door LAION van de website van het fotoagentschap gedownloade – van een watermerk voorziene – kopie van de foto was naar het oordeel van de rechter vrij op het internet beschikbaar, zodat voldaan is aan de eis van “rechtmatige toegang” (art. 4 lid 1 DSM-RI) (r.o. 86-87). Hierna onderzoekt de Hamburgse rechter of er sprake is van een rechtsgeldige opt-out. Volgens art. 4 lid 3 DSM-RI is de TDM-beperking van toepassing “op voorwaarde dat het gebruik van de in dat lid bedoelde werken en andere materialen door de rechthebbenden ervan niet op passende wijze uitdrukkelijk is voorbehouden, zoals machinaal leesbare middelen in het geval van online ter beschikking gestelde content.”

⁴ Aldus T.W. Dornis, ‘The Training of Generative AI Is Not Text and Data Mining’, *EIPR* 2025/2, beschikbaar op SSRN: <https://ssrn.com/abstract=4993782>.

Op de website van het fotoagentschap was als onderdeel van de gebruiksvoorwaarden de volgende tekst te lezen: *RESTRICTIONS. YOU MAY NOT [...] Use automated programs, applets, bots or the like to access the [...] website or any content thereon for any purpose, including, by way of example only, downloading Content, indexing, scraping or caching any content on the website.* De rechtsbank gaat uitvoerig in op de vraag of deze gebruiksbeperking als geldig voorbehoud in de zin van art. 4 lid 3 DSM-RI kan worden beschouwd. Allereerst de voorvraag of het voorbehoud door een ‘rechthebbende’ is gemaakt. Weliswaar heeft auteursrechthebbende Kneschke het voorbehoud niet zelf ingeroepen, maar als licentiehouder van Kneschke kwalificeert het agentschap als ‘rechthebbende’, zo meent het Landgericht (r.o. 90). Hoewel in het onderhavige geval het agentschap ongetwijfeld mede in het belang van Kneschke handelde, valt op deze overweging wel wat af te dingen. De term ‘rechthebbende’ wordt in de DSM-richtlijn weliswaar niet gedefinieerd, maar wel veelvuldig in de tekst en overwegingen gebruikt. Daaruit valt af te leiden dat de uniewetgever hierbij het oog heeft op houders van auteursrechten en naburige rechten, doch niet op licentienemers. Bovendien: het gaat niet aan dat een licentienemer een opt-out inroeft tegen de wil van de maker of auteursrechthebbende. Ik meen dat een licentienemer enkel gerechtigd zou moeten zijn een voorbehoud te maken indien deze daartoe door de auteur of (echte) rechthebbende is gemachtigd.

Vervolgens de hamvraag: is het voorbehoud met “machinaal leesbare middelen” gemaakt? Blijkens de Duitse wetsgeschiedenis is de eis van machineleesbaarheid bedoeld om het automatisch ‘crawlen’ van websites mogelijk te maken. ‘Machine-leesbaar’ moet daarom worden uitgelegd als ‘machine-begrijpelijk’ (r.o. 96), zo meent het Landgericht. Afhankelijk van de stand van de techniek zou ook een in natuurlijke taal gesteld voorbehoud rechtsgeldig kunnen zijn, zo meent de rechter. Daarbij verwijst de rechtsbank naar de in art. 53 lid 1 sub c van de AI Act aan ontwikkelaars van AI-modellen opgelegde verplichting om op grond van art. 4 lid 3 DSM-RI gemaakte opt-outs te herkennen en na te leven “onder meer door middel van geavanceerde technologieën”. Daartoe behoren, zo overweegt het Landgericht, AI-toepassingen die in staat zijn in natuurlijke taal gestelde teksten te begrijpen.

Ook dit onderdeel van het vonnis overtuigt mij niet. De Hamburgse rechter heeft gelijk dat een geavanceerd AI-taalmodel een in natuurlijke taal gestelde juridische tekst zou kunnen interpreteren. Maar de ‘bots’ waarmee websites op grote schaal worden gecrawled zijn zeer eenvoudige programma’s die niet zijn uitgerust met kunstmatige intelligentie. Zoals blijkt uit het ontwerp van de *AI Code of Practice*, de gedragscode die onder leiding van de AI Office van de Europese Commissie per april 2025 zijn beslag zal

krijgen, is het prototype van de machine-leesbare opt-out een volgens het Robots Exclusion Protocol (*robots.txt*) geprogrammeerde machine-leesbare instructie die door de gangbare bots automatisch herkend en begrepen kan worden.⁵ Daarnaast committeren de ondertekenaars van de Code zich ertoe “to making best efforts that are proportionate to their size and capacities in accordance with widely used industry standards to identify and comply with, including through state-of-the-art technologies, other machine-readable means to appropriately express a rights reserva-

tion at source or work level pursuant to Article 4(3) of Directive (EU) 2019/790 [...].”⁶ Nergens in de Code wordt gesproken van crawlers die opt-outs in natuurlijke taal zouden kunnen of moeten herkennen. De opvatting van de Duitse rechter staat daarmee ver af van de zich ontwikkelende praktijk van de marktdeelnemers.

Hoe dan ook is over de juiste uitleg van art. 4 DSM-RI het laatste woord nog lang niet gesproken. Inmiddels heeft Kneschke hoger beroep aangetekend.