

# **A mesterséges intelligencia etikai kérdései a radiológiában**

## **Európai és Észak-Amerikai radiológiai társaságok és egyéb szakmai szervezetek közös közleménye (előzetes anyag)**

Eredeti cím: „Ethics of AI in Radiology: European and North American Multisociety Statement”  
([https://www.acrdsi.org/-/media/DSI/Files/PDFs/Ethics-of-AI-in-Radiology-Statement\\_RFC.pdf?la=en](https://www.acrdsi.org/-/media/DSI/Files/PDFs/Ethics-of-AI-in-Radiology-Statement_RFC.pdf?la=en))

Informális fordítás radiológus szakmai közegben történő megvitatáshoz.  
Készítette: Bódi Péter szervizmérnök, Pécsi Diagnosztikai Központ  
Pécs, 2019. március 17.

Összefoglaló	3
Bevezető	4
Jelen közleményről	6
Adatok etikai kérdései	7
Klinikai radiológiai adatok	8
Gazdasági és analitikus adatok	8
Elő-tanuló, szintetikus és bővített adatok	9
Nyers képi adatok	10
Adatok tulajdonjoga	11
Adatmegosztás és adathasználat	12
Adatok bizalmassága	15
Adatmódosulás és torzulás	17
Adatok jelölése és alapigazság	19
Algoritmusok és tanuló modellek etikája	21
Gépi döntéshozatal	21
Algoritmus-választás	22
Algoritmus-tanítás	23
Modellértékelés és tesztelés	23
Átláthatóság, értelmezhetőség és indokolhatóság	25
Megismételhetőség	26
Algoritmus adattorzulás	26
A gyakorlat etikája	27
Számítógép és emberi interakció – a döntési körben maradó humán tényező	27
Oktatás	28
Automatizálási elfogultság	29
Páciens preferenciák	30
Nyomonkövethetőség	30
Mesterséges intelligencia és munkaerő-erózió	31
Erőforrás egyenlőtlenség	32
Felelősség	32
Érdekkonfliktusok	33
Következtetések	34
Meghatározások	35
Irodalmi referenciák	36

## Összefoglaló

A mesterséges intelligencia (AI; Artificial intelligence) olyan számítógépes alkalmazásokra vonatkozó fogalom, amelyek olyan módon viselkednek, amiről a legutóbbi időig úgy gondoltuk: csak emberi elme képes ilyen működésre. Az AI hatalmas lehetőség, mely a radiológia minden frontján alapvető előrelépést és fejlődést hozhat [1]. Az AI komplex kérdéskör, számos potenciális buktatóval, és bizonyos fokig befolyásolt és előítéletes vélekedések is kapcsolódnak hozzá. A radiológusok - és mindazok a szakemberek, akik létrehozták és használják az AI produktumokat - kötelesek mélyebb ismereteket szerezni az AI témakörben. Ennek révén a páciensek számára optimális előnyöket nyújthatnak, megérthetik: mikor és hogyan lépnek fel AI-hez kapcsolódó veszélyek. Ugyanakkor képesek lesznek átlátni és mérlegelni a kockázatok és előnyök viszonyát, és a kockázatokat a lehető leghatékonyabban mérsékelni. Nem kétséges: az AI drámai változásokat hoz a klinikum, a társadalom és a gazdaság területén egyaránt. A legtöbb változás pozitív lesz, de némely hatás negatív is lehet.

Az AI jelentősen megváltoztatta a radiológiai adatokkal kapcsolatos felfogásunkat: mi is ezeknek az adatoknak a valódi értéke, hogyan használjuk az adatokat, mik a téves vagy rosszindulatú adathasználat veszélyei. Az AI módszer egyszerű alkalmazásán túl az általa gyűjtött és kezelt adatok megértése is a radiológus erkölcsi kötelessége. Ezáltal a „közjó” érdekében több információt nyerhetnek ki az adatokból a páciensekről és betegségeikről, így hozzájárulnak a radiológia gyakorlatának jobbításához.

Az adatok torzulása, befolyásoltsága - vagyis a valósághoz képesti szisztémás eltérése vagy torzulása (bias) bizonyos mértékig minden adatkészletnél előfordul. Ez különböző módokon nyilvánulhat meg, ezek mindegyike alapos elemzést és tudatos figyelmet érdemel, hogy az adattorzulás hatása az AI modellek révén létrejött döntésekre minimális legyen.

A radiológiának már ma el kell kezdeni az AI számára szóló etikai szabályok és eljárásrend kidolgozását. Ezeknek a szabványoknak és szabályozásnak, az etikus AI alkalmazásokra vonatkozó eljárásrendnek a műszaki, klinikai és gazdasági érdekek bonyolult szövevényében kell eligazítással szolgálni, mégpedig kellő morális kritikával. Az etikus AI létrehozása és biztosítása minden résztvevő részéről maximális bizalomra törekvést követel meg. Radiológia-centrikus AI szaktudás és etikus technológia egyaránt szükséges az AI produktumok ellenőrzéséhez és engedélyezéséhez. A verifikáláshoz és validáláshoz szükséges eljárásrend kulcsfontosságú elemei a folyamatos és hangsúlyos transzparencia, a páciensek védelme, az adatkészlet változatok gondos kézbentartása és alkalmazása. Az AI alkalmazások folyamatos és alapos felügyeletet és ellenőrzést igényelnek ahhoz, hogy az elvárásoknak megfelelően működjenek, hogy az általuk hozott döntések optimális és etikus gyógyító tevékenységet eredményezzenek.

A világ radiológus társadalma folyamatosan hall és tanul az AI módszerekről, és ezzel párhuzamosan próbálja kialakítani, bevezetni és használni ezt az új technológia elemet. Korunkban mindez alig felfogható sebességű és kihatású technológiai fejlődés közepette zajlik. Az AI jelentősen befolyásolja majd a radiológus szerepkörét és pozícióját, forradalmasítja a radiológiai leletekhez kapcsolódó döntéshozatalt, és erőteljesen

megváltoztatja a radiológus-páciens kapcsolatot, valamint a radiológus és az egyéb egészségügyi szereplők viszonyát.

## Bevezető

A jelen dokumentum a nemzetközi radiológus társadalomnak a radiológiában alkalmazott AI (Artificial Intelligence – mesterséges intelligencia) technológiához kapcsolódó etikai elvekre és viselkedési szabályokra vonatkozó elemzési igényét fejezi ki. Célunk, hogy a folyamatban résztvevők kölcsönös bizalmát megnyerjük: az AI jó szolgálatot tesz a pácienseknek és a radiológus társadalomnak, valamint hogy áttekintsük: miként érvényesülnek etikai szempontok és szándékok az AI radiológiában való alkalmazásával kapcsolatban. Az itt felsorolt témákban való kutatást ösztönzendő, körvonalazzuk az autonóm és intelligens radiológiai rendszerek tervezésével és alkalmazásával kapcsolatos etikai kérdéseket. A páciensek érdekében segítünk megérteni az ilyen rendszerek működési mechanizmusát, és azt is: miként kerülhető el, hogy veszélyesek vagy ártalmasak legyenek a páciensekre nézve. Mindemellett megvizsgáljuk ennek a szakterületnek a szabályozásával, az alkalmazandó eljárásrenddel kapcsolatos szempontokat, továbbá szemléltetjük azokat az orvosi, kulturális és gazdasági tényezőket, melyek a radiológia, az etika és az AI eszközök kölcsönhatásánál érvényesülnek.

A radiológusok több, mint egy évtizedes orvosi oktatás, specializált képzés során tesznek szert olyan szaktudásra, mely szükséges a radiológiai képek elemzéséhez. Ennek révén képesek lesznek részletes és mélyreható, gyakran az életet vagy életminőséget meghatározó információk kinyerésére az emberi szervezetben zajló folyamatokról. A páciensek, az orvosi környezet és a társadalom egyaránt azt várják a radiológusoktól, hogy a képadatok alapján döntéseket hozzanak. Ez a speciális döntéshozatali képesség és elvárás bizonyos hierarchiát határoz meg a radiológusok és döntéseikre számítók között. A radiológusok szakmai etikai szabályrendszere biztosítja, hogy szaktekintélyük és tevékenységük morálisan vállalható, pozitív eredményeket hozzon. Az AI és a Machine Learning (ML, adaptív rendszerek gépi tanulása) **alapvetően statisztikai eljárások összességét jelentik**, mint ilyenek segítik a radiológust a radiológiai vizsgálatokból kinyerhető adatok mennyisége tekintetében, gazdagítják a radiológia döntési eszköztárát és javítják a páciensek ellátását.

A haladással az orvosi képadatokból levont következtetések létrehozói nem csupán humán szakemberek lesznek, hanem a folyamatba várhatóan bekapcsolódnak majd intelligens gépi eljárások is. Esetenként a gépi döntések jobbak lehetnek, lehet, hogy a gépek gyorsabbnak és hatékonyabbnak bizonyulnak a humán radiológusoknál, esetleg döntéseikben nekik ellentmondhatnak. Az AI befolyásolja a képi adatok értelmezését, és azt is, hogy mi és milyen módon kerüljön az orvosi leletbe. Ennek során megváltozik az eredmények kommunikációja, de módosulhat a radiológiai szolgáltatásokért felszámított díj is. Az AI eszközök képesek a professzionális viszonyok befolyásolására, a páciens bevonásának megváltoztatására, a szaktudás hierarchiájának módosítására, és a munkaerőpiac átrendezésére is. Ráadásul az AI révén fokozódhat az erőforrások koncentrációja és kiegyenlítetlensége is. Ennek eredményeként jelentős AI eszköztárral rendelkező, több és jobb radiológiai döntéshozatali lehetőséggel rendelkező intézmények jöhetnek létre. Ebben a folyamatban maguk a radiológusok és munkahelyeik is „adattá” válnak, az alkalmazott AI modellek kategorizálják,

vagyis besorolják majd a radiológusokat és intézményeiket. Az AI specifikus viselkedési mintázatokat eredményez majd a személyes, szakmai és intézményi szinten egyaránt. Megváltoztatja hagyományos véleményünket arról, hogy mennyire valóságűek és etikusak a radiológiai adatok. Ugyanígy módosul majd a radiológiai adatok birtoklásának, hozzáféréseinek módja, végeredményben pedig az is: ki és mit tud majd, és hogyan alkalmazza ezt a tudást.

Miközben az AI eszközök a leletezés minőségének és hatékonyságának javulását, a költségek csökkenését ígérik, egyben új lehetőségeket és következményeket eredményeznek. Új kérdéseket vetnek fel a páciens és a radiológus társadalom számára egyaránt. A felmerülő kérdéseket egyrészt a társadalom etikája, de ugyanilyen mértékben technikai tényezők is alakítják. Más hatások kevésbé közvetlenek: ilyenek az intézményi vagy a társadalmi eljárásrend algoritmusai, melyek döntéseket befolyásolnak, vagy összefüggéseket keresnek a nagy populációs adathalmazokban, hogy az eredményekkel javulást hozzanak az egészségügyi közellátásban, a betegségek megértésében és gyógyításában.

A radiológia kötelessége az AI módszerek aktív követése és alkalmazása saját tevékenységének javítása érdekében. Ugyanakkor figyelemmel kell kövesse ezt az adatfeldolgozó, ember-gép együtttest alkalmazó döntéshozatali mechanizmust, hogy a jövőben megelőzze a nemkívánatos, optimális betegellátástól eltérítő hatásokat és következményeket. Rendszeresen és gyors ütemben újabb és újabb etikai kérdések merülnek fel, és társadalmunk ezekkel kapcsolatos felfogása is nyilván változik majd. Míg jelenleg is fontos a radiológiában alkalmazott AI etikai kérdéseinek elemzése, ugyanilyen lényeges, hogy ezt a témát a jövőben – az okozott hatásokról szerzett tapasztalataink bővülésével párhuzamosan – újra és újra vizsgálat tárgyává tegyük.

Kezdetben a radiológiában alkalmazott AI jórészt intelligens, a radiológus munkafolyamatába illeszkedő, klinikai döntéshozatali támogató modelleket jelent majd. Ilyenek például a már napjainkban is alkalmazott mérési eszközök vagy a számítógéppel támogatott észlelés (CAD; Computer Assisted Detection). A jelek szerint az AI rendszerek fokozatosan egyre autonómbbabbak lesznek, önállóan hozhatnak döntéseket és kezdeményezhetnek műveleteket, alkalmasint a radiológus felügyelete nélkül is.

Más iparágak jelenségeiből levonható tapasztalatok alapján, a jövőbe tekintve megállapítható: az AI eszközökkel ellátott radiológia komplex, dinamikus hálózatba szervezett környezeté fejlődik majd [4]. A szofisztikált, hálózatba szervezett autonóm algoritmusok szakosodott radiológusok csapatához hasonlíthatók, ahol minden egyes résztvevő a diagnosztikai probléma egy adott részletére fókuszálva hoz döntéseket. Attól függően, hogy véleményükkel egyezsége jutnak-e, a feldolgozott vizsgálatot további autonóm algoritmusoknak adják tovább. Ez utóbbiak, saját hatáskörükben szintén döntéseket hoznak a diagnózis adott részével kapcsolatban. Az ilyen kaszkadosított, idővel akár struktúráját is változtató hálózatban szervezett döntéshozó számítógépek komplex együttese hozza majd létre, továbbítja egymásnak és fogadja el más komponensek információját.

A fentebb leírt dinamikus önszervező hálózatos rendszerek eddig még nem nagyon léteztek – évek telnek el, mire megtervezik és létrehozzák őket. Még sokat kell tanulnunk, hogyan is kellene ilyen rendszert robusztus, biztonságos, pontos és megbízható módon megszervezni.

Nemkülönben sok munkára lesz szükség az így létrejövő rendszerek viselkedését és belső működési logikájának megértéséhez.

A radiológusok, akik végső soron felelősek maradnak azért, ami a páciensekkel történik, képessé kell váljanak az így kialakuló döntési ökoszisztémák menedzselésére, folyamatosan szem előtt tartva a páciensek érdekeit és javát. A radiológus társadalomnak szüksége van a technológiai fejlődést kézben tartó etikai vezérelemekre is. Egyben képesnek kell lenniük az egészségügyi beruházókat befolyásolni az AI elfogadásával és alkalmazásával kapcsolatban, ügyelve arra, hogy ezen eszközök alkalmazásával kapcsolatos döntéseik egyre jobban szolgálják a páciensek érdekeit – lehetőleg magukat a pácienseket is bevonva a folyamatba.

Mivel egyes AI modellek létrehozása és betanítása aránylag egyszerű feladat, fennáll a veszély, hogy a termékfejlesztésben és az AI alkalmazások kereskedelmi forgalmazásában nagyszámú naiv és nem kellően szakszerű szereplő is felléphet. Mindez fokozza az orvostudományban, statisztikatudományban és számítástechnikában meglévő etikai szabályrendszerek kibővítésének jelentőségét, különös tekintettel a radiológiában alkalmazott AI rendszerek döntési sajátosságaira [5–7].

Számos, az orvostudományon és orvosi szakmai szervezeteken kívüli szakterület elemzi az AI etikai kérdéskörét. Legutóbb a New England Journal of Medicine (NEJM) és a Journal of the American Medical Association (JAMA) cikkei írtak az AI ígéretes kilátásairól [8], és az AI alkalmazásokkal kapcsolatos adattorzulás, adateltérés (bias) potenciális problémáiról, valamint a páciensek és az AI bizalmi viszonyának kérdéséről [9, 10]. A mérnöki és informatikai tudományok vezető szakmai fórumai, köztük az Institute of Electrical and Electronics Engineers (IEEE), az Association for Computing Machinery (ACM), Future of Life Institute, valamint kormányzati szervek, mint az European Commission's Group on Ethics in Science and New Technologies (Új technológiák és tudomány etikai kérdéseivel foglalkozó EU Bizottsági csoport) folyamatosan frissítik az AI-vel kapcsolatos ajánlásaikat és vezéreleiket [11–14].

## Jelen közleményről

Jelen közlemény több szervezet, az American College of Radiology (ACR), European Society of Radiology (ECR), Radiology Society of North America (RSNA), Society for Imaging Informatics in Medicine, European Society of Medical Imaging Informatics, Canadian Association of Radiologists, és az American Association of Physicists in Medicine közös erőfeszítésével jött létre. A szerkesztőbizottság magját képező csoportban volt amerikai filozófus, voltak észak-amerikai és európai radiológusok, képző specialista informatikusok, orvostudományi fizikus, páciens érdekvédelmi szakemberek és az EU és USA radiológiai-jogi kérdéseiben jártas ügyészek is.

Ezt az előzetes közleményt a támogató és létrehozó társaságok egyike sem jegyzi specifikusan. Mi, a szerkesztőbizottság azzal a felhívással hozzuk nyilvánosságra közleményt: az összes érdekelt fél küldjön be véleményeket, hozzászólásokat akár a közleménnyel, akár a radiológia etikai kérdéseivel kapcsolatban. Különösen bátorítjuk az AI

technológia által érintett páciensek és egyéb szereplők véleménynyilvánítását és kommentárjait. Szívesen fogadnánk, és nagyra értékelnénk az etika, jog és adattudományok területén jártas, az adott terület iránt határozottan érdeklődő szakemberek véleményét – várjuk az észrevételeket, megjegyzéseket és javaslatokat. A beérkező hozzászólások alapján véglegesített közleményt adunk majd ki, a véleménybeküldésre meghatározott 2019. április 7.-i határidőt követően mintegy 6 hónap múlva.

Ennek a közleménynek szerkesztése során áttekintettük a számítástechnikai- és orvostudomány jelenleg fellelhető etikai szakirodalmát, az etikai kérdések történelmi iskoláinak anyagát, valamint a jövő etikai helyzeteinek kérdéskörét is. A hatékonyság érdekében vizsgálódásunk - és így közleményünk is - az észak-amerikai és európai helyzetre összpontosít. Ugyanakkor látjuk, hogy más régióknak eltérő, vagy esetleg részben átfedő értékrendszere és etikája lehet.

Jelen közleményt ösztönzőnek és inspiratívnak szánják a szerzők, nem előírásnak. Célunk az AI etikájára vonatkozó, aránylag könnyen megérthető és alkalmazható megközelítés közzététele. Várakozásaink szerint a témakör a technológia és tudomány előrehaladásával rohamosan fejlődik majd, melynek során az automatizált döntéshozatalt kezelő számos új jogi megközelítés és felelősségi meghatározás fejlődik ki. Például Kalifornia Állam új adatvédelmi törvénye [15, 16] és az EU GDPR [17] szabályozása, vagy az Európai Parlamentnél javaslati szakaszban lévő Civil Law Rules on Robotics [18] egy ilyen törvényhozási folyamat előhírnökeinek tekinthetők. A kereskedelmi forgalomba kerülő, általános radiológiai alkalmazásra szánt AI termékek előállítóinak támogató és meghatározó etikai útmutatásra van szüksége. Jelen közlemény az ilyen szabályozás megszerkesztésével foglalkozó felügyeleti szervek jövőbeli munkáját segíti. Ebben az előzetes közleményben nem sok gyakorlati útmutatással szolgálunk, de szándékunk szerint a végleges közlemény több praktikus javaslatot tartalmaz majd.

Az AI radiológiában történő etikus használata során tisztelni kell az emberiség alapvető etikai elveit, szem előtt kell tartani a szubjektumnak a biomedikai és viselkedéstudományi kutatásokkal szembeni védelmét. A radiológia etikai kérdései közül némelyek igen messze vezetők és mélyrehatóak. Ilyen esetekben megpróbálunk óvatosságra inteni azokkal a problémákkal kapcsolatban, melyek etikailag különösen szorítóak. Megmagyarázzuk, hogy az egyes etikai problémák hogyan hatnak a radiológiában, ugyanakkor javasolunk olyan szempontokat, melyeket a szakmai közeg számításba kell vevni. Ahol specifikusan a radiológiával kapcsolatos problémát találunk, ott megoldási stratégiát javasolunk, amennyiben a kérdésre elegendően világos válasz és megoldás adható.

Jelen közlemény szerkezetében az [5] szakirodalomban Floridi és társai által javasolt irányelveket követi. Elemzésünkben az etika tárgykört adatok etikája, algoritmusok etikai kérdései és a gyakorlat etikája szakaszokra osztjuk.

## Adatok etikai kérdései

Az adatok etikai kérdései alapvetők az AI radiológiai alkalmazásában. Kulcsfontosságú ebben a körben az ún. megfelelő tájékozottságon alapuló beleegyezés kérdése, az adatok

bizalmassága és védelme, az adatkészletek eltérése vagy torzulása, valóságghűsége, tulajdonjoga, objektivitása, az adatkezelés átláthatósága. Ugyanilyen lényeges az a probléma, hogy míg egyes szereplők rendelkeznek a nagy adatbázisokhoz való hozzáférés jogával és eszközeivel, mások azonban nem. További adatokkal kapcsolatos kérdéses problémák a nem, az etnikum és a gazdasági helyzet által meghatározott csoportszintű adatkészletek eltérései vagy torzulásai, az adatetikai kérdéskörben a bizalom, vagy az, hogyan biztosítsunk értelmesen megvalósítható és morálisan helyes hozzáférést az adatokhoz [6].

Az AI alapvetően megváltoztatta felfogásunkat a radiológiai vizsgálatokról és az ahhoz kapcsolódó adatokról, az adatok alkalmazásáról, de az adatokkal kapcsolatos esetleges visszaélésekről is. A radiológus kötelessége felfogni az AI alkalmazás problematikáját, de ugyanígy kötelezett megérteni az adatkezeléssel kapcsolatos kérdéseket. A szélesebb értelemben vett radiológus társadalom erkölcsi kötelessége az összegyűjtött adatok közjó érdekében történő felhasználása, az adatokból minél több információ kinyerése a páciensről és betegségéről, összességében a radiológia gyakorlatának jobbítása. Etikai kötelességük a páciensektől származó adatok minél hasznosabb feldolgozása éppen azok érdekében, akiktől és akikről az adatokat nyerték.

## Klinikai radiológiai adatok

A képkötő vizsgálatok eredménye tipikusan egy adatkészlet és ahhoz szorosan kapcsolódó jelölők (labels) összessége [20].

A primer képadatokat vizsgálóberendezések állítják elő, ezt követően képfeldolgozás alakítja azokat emberi megfigyelő számára megtekinthető és értelmezhető formátumú képekké. A primer ún. nyers-adatok közvetlenül nem értelmezhetők a radiológus számára, a feldolgozás során képelemekké, pixelek rendezett halmazává alakulnak át – ezt nevezzük radiológiai képnek. Pixelek mátrixa alkotja a humán értékelő által elemzett képeket. Míg a radiológiai vizsgálat után a pixelek alkotta radiológiai képek - jellemzően további kiegészítő meta-adatokkal együtt – tárolásra kerülnek, a nyers-adatok általában rövid idő után (pl. 72 óra elteltével) törlődnek a képkötő modalitás számítógépéből. Bizonyos esetekben, mint pl. ultrahang képkötésnél, a meta-adatok (például páciens információ) beágyazhatók a pixeladatok közé. Az ilyen eljárást beépülő vagy „beégo” adatoknak nevezik. Jelenleg a legtöbb képalapú AI eljárás pixeladatokot használ, de vannak törekvések a nyers-adatok feldolgozására is, melyek sok esetben a pixel-képadatoknál jelentősen több információt hordoznak [8].

A jelölők a képhez tartozó további adatok, információk és mérési értékek. Ezek a jelölők lehetnek vizsgálat-szintű leírók (mint pl. „abdomen MRI”) vagy kép szintű attributumok (pl. 36. kép; máj pixeles ábrázolása). A képeket kísérő, szöveges radiológiai leletek a képadatokból kinyert észleléseket, értelmezéseket és diagnózist tartalmaznak, mindezek rendszerint a jelölők forrásaiként szolgálnak. Jelölők közé sorolhatók az alábbiak:

- Radiológiai lelet észlelések, melyekben „közös adatelemek” (CDE, Common Data



Element) is szerepelnek

- Kép annotációk vagy feliratok, lehetnek grafikus elemek pl. nyíl ábra vagy tartománykijelölés (RoI, Range of Interest) vagy mérési értékek
- Specifikusan az AI feldolgozás számára hozzáadott jelölők
- Nem kép jellegű klinikai adatok, az elektronikus egészségügyi adatlapról (EHR, Electronic Health Record) származó dokumentáció, patológiai, laboratóriumi vagy genetikai leleteredmény stb.
- Szociális médiából származó vagy egyéb nyilvánosan hozzáférhető adat, mint pl. időjárás vagy geográfiai térképadat
- Egyéb adatok, melyet vagy a páciens, a nyilvánosság, esetleg egyéb internetes forrás (IoT; Internet of Things) szolgáltat

A képalapú AI rendszerek teljesítőképessége a pixeladatok diverzitásától, és a képadatokhoz tartozó jelölők pontosságától, részletességétől függ. A radiológia javíthatja az AI rendszerek minőségét az annotációk és mérések egységesítésével, az adatok és adatváltozatok nyomon követésének szabványosításával, adatmódosító, adattovábbító eljárás standardok bevezetésével, továbbá a páciens eredmények és a hozzátartozó meta-adatok korrelációjának megállapításával és jegyzésével [20].

## Gazdasági és analitikus adatok

Gazdasági működési adatok közé tartoznak a felhasználó tranzakciós adatai, az alkalmazottak munkafeladatai, üzleti folyamatok. Az IT (Information Technology) működési adatai leírják, milyen adatokat és milyen hatékonysággal kezel az információ-technológiai rendszer. Az üzleti és IT analitikus adatok az informatikai folyamatok sebességét és pontosságát írják le, tájékoztatnak az üzleti-technológiai rendszer kockázatairól és biztonságáról, továbbá informálnak az adatok épségével, integritásával és validálásával kapcsolatos eljárásokról, az üzleti hatékonyságról és termelékenységről. A radiológiai leletek elkészülésének ideje, a teljesítmény elszámolását szolgáló relatív egységek (RVUs, Relative Value Units), a vizsgálóberendezés kihasználtsága, a képalkotó eljárások minőségi értékelése mind tipikus példák a klinikai radiológiában szereplő gazdasági működési adatokra.

## Elő-tanuló, szintetikus és bővített adatok

Az AI modellek teljesítőképessége javul, ahogy a rendszer egyre több és több adaton „tanul”. Az orvosi képfeldolgozáson kívül eső szakterületekről érkező hírek azt mutatják, hogy a több millió vagy akár több tízmillió képet tartalmazó adatkészletet feldolgozó, képfelismerő AI modellek elképesztő hatékonyságúak lehetnek. Ezzel összehasonlítva a jelenlegi radiológiai AI rendszerek számára rendelkezésre álló adatbázisok tipikusan néhány száztól néhány tízezerig terjedő számú radiológiai vizsgálati mintaelemet tartalmaznak. Ennek eredményeként a mai radiológiai AI modellekben működő algoritmusok lehetnek hatalmas, nem-orvosi jellegű képi adatbázison (ilyen az ImageNet; ez több mint 14 millió, általános objektumot, mint pl. kutyák, autók vagy tájak képeit tartalmazó adatbázis) „trenírozott”, elő-

tanuló eljárások. Más módszerek szintetikus, vagy ún. augmentált adatokat is alkalmaznak a tanulási folyamathoz [22, 23]. Egy adott típusú adatkészleten trenírozott algoritmus-modellek alkalmazását egy másik típusú adatkészletre nevezzük „transzfer-tanulásnak”.

A trenírozáshoz rendelkezésre álló adatkészlet bővítésének egyik lehetséges módszere teljesen, vagy részben mesterségesen létrehozott adatok alkalmazása, melyeket általában szintetikus adatnak nevezünk. A szintetikus adatokat - legalábbis részben - statisztikai programok hozzák létre, melynek során az adatok véletlenszerű, random tulajdonságúak lesznek. Ha az ilyen mesterséges adatokat előállító modellt kifejlesztették, ettől kezdve a szintetikus adatok generálása egyszerű, gyors és olcsó. Az ilyen szintetikus adatok hasznosak az elő-trenírozó (pre-training) folyamatban [24]. Nem áll fenn a veszélye annak, hogy páciensadatok kerülnek a szintetikus adatok közé, mivel az adatok nem valós, élő alanyoktól származnak. A radiológiában az ilyen szintetikus adatok képesek lehetnek ritka betegségek szimulálására, ami az algoritmusok számára lehetővé teszi a tanulást jellemző patológiát mutató képeken olyan esetekben is, amikor ilyen képadatok beszerzése valós klinikai páciensektől nehézségekbe ütközne. Hasznosak a kutatók és fejlesztők számára is, amikor egyébként nem állna rendelkezésre kiinduló adat. Alkalmosak továbbá adatok előállítására AI rendszerek tesztelése vagy verifikálása számára. Szintetikus adatokat gyakran alkalmaznak ún. „adverzariális” képekként, az AI algoritmusok egy osztályát képező adverzariális (szembenálló, vetélkedő feladatmegoldó módszer) hálózatokban [25].

Augmentált képadatok olyan valós forrású adatok, melyeknél adatmásolás során minden másolat valamely tulajdonságban különbözik a kiinduló adattól [26]. Az augmentáció általában képforgatást, döntést, tükrözést vagy átméretezést jelent, de szerepelhet hozzáadott zaj, képfélesztés is a transzformációs folyamatban. Augmentált trenírozó adatok hasznosak lehetnek például olyan esetben, ha cél az, hogy egy AI rendszer a változások ellenére felismerjen egy adott objektumot. Sok esetben az ilyen augmentált adatok előállítása könnyebb, mint a szintetikus adatoké, bár a valós páciensképekből származó augmentált adatok esetén fennállhatnak személyiségi jogi vagy adatvédelmi korlátozások.

A szintetikus és augmentált adatok segítségünkre vannak a valódi képadatokban hiányzó kiegészítések pótlására és kiegészítésére. Hasznosak a leletezés és képkiválasztás során fellépő adattorzulás és adateltérés (bias) javítására. Ugyanakkor esetleg felnagyíthatják és eltúlozhatják az adateltérést azzal, hogy duplikálják vagy megerősítik [27] a forrásul szolgáló, kiinduló adatbázisban meglévő inherens adateltérést. Bár világos, hogy az ilyen adatok hasznosak lehetnek a tanuló algoritmusok trenírozásánál, jelentős további kutatás szükséges a lehetséges elágazások megértéséhez, továbbá ahhoz, hogy meghatározható legyen: hol a kiterjedt mesterséges adatbázisok alkalmazásának határa a radiológiában, és milyen kritériumok kellenek az ilyen adatbázisok alkalmazásához.

## Nyers képadatok

Nyers-adatok rendszerint a vizsgálóberendezéseket (pl. CT vagy MRI szkennerek) gyártó cégek saját tulajdonú és formátumú adatai. Ezek az adatok jórészt értelmezhetetlenek az ember számára. A digitális radiológia kezdeteinél a digitális adattárolás költséges dolog volt. Ezért a berendezések eleinte csak a klinikailag relevánsnak tartott formátumú képadatokat mentették

el, míg a nyers-adatokat ritkán tárolták egy adott rövid, a képelőállításához és feldolgozáshoz szükséges intervallumnál hosszabb ideig. Elméletileg fennáll a lehetőség, hogy az AI modellek a nyers-adatokból eredményesebben és megbízhatóbban tudnak kérdéses sajátosságokat megállapítani, mint az emberi értékelő számára értelmezhető képadatokból. Ezért a radiológia egyre inkább felismeri a nyers-adatok jelentőségét. A páciensek, a kutatók és a diagnosztikai ipar is nyerne azzal, ha a hagyományos leletezésre szánt, utófeldolgozott képi adatok mellett a nyers-adatokat is elmentené, tárolná és elérhetővé tenné a vizsgálóberendezés [20].

## Adatok tulajdonjoga

Az egészségügyi szolgáltató egységek gyűjtik, és védetten tárolják a páciensek orvosi képeit és a kapcsolódó orvosi információt. Mostanában, ahogy erőteljes és hatékony módszerek alakultak ki az adatok továbbítására és megosztására, és igény van az adatok tömeges kigyűjtésére AI módszerek fejlesztéséhez. A képadatok megosztott használata egyre általánosabb a radiológusok, egyéb adatkezelők, intézmények vagy akár egyes országok között is. Az adatok biztonságos kezelésének etikai és technikai kérdései bonyolultak, melyet tovább nehezít az, hogy az erre vonatkozó normák országonként eltérőek. Ez a komplexitás és eltérés akadályozza a páciensadatok megosztását a klinikai ellátásban, az AI kutatásban és jelentős akadályt képez a gazdasági fejlődésben is.

Annak a kérdésnek eldöntésében, hogy „kié a páciensadat?“, a radiológus, a szélesebb medikai környezet, és a szabályozást képviselő hatóságok egyaránt érintettek. Az adat-tulajdonlás kérdésének szabályozása országonként eltérő. Az Amerikai Egyesült Államokban a képalkotást végző egység lesz az adat birtokosa, melyet a páciens adatmásolathoz való joga egészít ki. Miközben a gyakorlat vegyes képet mutat, sok kórház az általános kezelési beleegyező nyilatkozatba belefoglalja az adatok utólagos, kutatási célú felhasználási lehetőségét is. Ezt a gyakorlatot a páciensek a tapasztalatok szerint elfogadják [28]. Az Amerikai Egyesült Államokban a szövetségi törvénykezés nem követeli meg az anonimizált adatok későbbi, kutatási célú tanulmányokban történő felhasználásához a hozzájárulást. Ezt a 45 CFR 46 (2018-as változat) USA szabályozás kivonata a következőképp határozza meg:

(ii) az olyan adatokat, melyek biológiai egyedektől származó információt tartalmazhatnak, a vizsgálónak olyan módon kell feljegyezni és tárolni, hogy az eredeti alany személyazonossága ne legyen meghatározható közvetlenül az adatokból vagy azokhoz kapcsolódó azonosítókból. Továbbá a kutató-vizsgáló nem veheti fel a kapcsolatot az alannyal, és a kutató-vizsgáló nem azonosíthatja ismét az alanyt...[19]

Ezzel szemben az EU adatokra vonatkozó GDPR (General Data Protection Regulation) szabályozása szerint a páciens birtokolja a saját érzékeny, személyes és/vagy azonosíthatóan hozzátartozó egyéb adatokat (melyek lehetnek orvosi és nem-orvosi adatok egyaránt), és rendelkezik azok felett. A GDPR egyértelmű és explicit páciens beleegyezést ír elő adatainak megosztásához vagy ismételt felhasználásához, mely beleegyezés bármely időpontban visszavonható [17]. Minden európai uniós országnak kell legyen az adatvédelemért felelős, nemzeti szakhatósága [29].

Egy újabb keletű európai kezdeményezés indítványozza, hogy a páciensek „ajándékozzák”

oda (név nélküli) adataikat a képalkotó vizsgálatot és diagnózist követően [30]. A radiológiai vizsgálatokat végző munkahelyekre is vonatkoznak a személyiségi és tulajdonjogokat védő szabályozások, melyből következően a radiológiai adatok további felhasználásához a páciensnek és az intézménynek egyaránt hozzájárulást kell adnia. Kanadában az USA-hoz hasonlóan, az orvosi képeket előállító egészségügyi szolgáltatók lesznek a fizikai regisztrátum tulajdonosai, és a páciensnek joga van saját adataihoz hozzáférni [31]. Kanadán belül az egészségügyi szolgáltatás szabályozása inkább tartományi mintsem szövetségi, területenként változó [32, 33]. Az ország legutóbbi „Tri-Council” eljárásrendje (Statement: Ethical Conduct for Research Involving Humans [34]) szerint: „nem szükséges beleegyezés az adatok olyan kutatási célú felhasználásához, mely kizárólag a nem azonosítható személyes információk másodlagos alkalmazásán alapul”. Ezt a szabályozást fogadják el Kanada legnagyobb kutatóintézményei, ami az AI-vel kapcsolatos kutatásokat lehetővé teszi majd.

Bár a személyes adatjogi és adat-tulajdonlási kérdések elemzése kívül esik ennek a közleménynek terjedelmén, megállapítható, hogy szükség van az adatok tulajdonjogának újabb tárgyalására, valamint annak elemzésére: az adatok továbbítása és újrahasználása esetén ki és mit birtokol, kinek és ki számára kell fizetni mindezért. Más szavakkal: lehet, hogy a képalkotó berendezés tulajdonosa birtokolja a pixeladatokat, míg a radiológus a tulajdonosa a tágabb értelemben vett jelölőknek, azaz az általa a életben létrehozott annotációknak, észleléseknek, konklúzióknak? A legutóbbi időkig, az orvosi adatok megosztása és koncentrációja leginkább a kutatást szolgálta, mely folyamatot kiforrott eljárásrend szabályozta. De ha az orvosi képalkotás adatainak értéke is kétkomponensű, azaz fizikai pixelek és leleteredmények együtteséből áll, akkor ez a közös érték jóval jelentősebb lehet, mint alkotóinak értéke külön-külön. Az még kérdés, hogy ennek a közös, nagyobb értéknek ki a tulajdonosa.

## Adatmegosztás és adathasználat

A keresőmotorok és szövegszerkesztők, a digitális asszisztensek korában az adatok értékének eltolódása és áttevődése egy üzleti modellt bomlasztott fel. Az említett hagyományos IT termékeket nem szolgáltatásnyújtásra, hanem inkább adatgyűjtésre, kapitalizációra és profittermelésre konstruálták. Ez a paradigma pedig elég hatékony és erős hatású ahhoz, hogy átterjedjen az orvostudományra, így a radiológiára is.

Ahogy az orvosi adatok egyre értékesebbek lesznek, elmosódik az adatok felhasználásának „akadémiai”, azaz tudományt szolgáló és a kereskedelmi, azaz profitáló arculata közötti határvonal. Példaként tételezzük fel, hogy egy kórház eladja képadatainak kizárólagos jogát egy IT vállalatnak abban a reményben, hogy az egy értékes AI terméket fejleszt ki. Mivel a páciensek joga saját adataikhoz megmarad, vajon a pácienseknek lehetőségük van-e arra, hogy ugyanezeket az adatokat egy másik vállalatnak adják el, ugyancsak AI termékfejlesztéshez? Vagy a páciensek - adatjogaik révén - elutasíthatják adataik kereskedelmi célú felhasználását, de esetleg engedélyezhetik azok non-profit, kutatási célú alkalmazását? Számos kormányzati és egyéb egészségügyi beruházó alap megköveteli a jelentkező páciensektől adataik megosztásának engedélyezését, vajon hogy lesz ez összeegyeztethető a kizárólagos adathasználati megállapodásokkal? A törvényhozó és

irányító testületeknek újra kell gondolni az orvosi adatok akadémiai-kutatási közegben és kereskedelmi területen történő alkalmazását szabályozó eljárásrendjüket. Ennek során méltányos egyensúlyt kell találni a társadalmi közjó szempontjai, valamint az adatokat szolgáltató egységek jogai és érdekei között [35].

A radiológiai adatok égbeszökő mennyisége szétfeszíti a hagyományos adatmegosztás kereteit, egyre gyakoribb jelenség a radiológia adatok adás-vétele. Az orvosi adatok kereskedelmének új üzleti szokásait befolyásolhatja a naivitás és a kapzsiság egyaránt. Például az Egyesült Királyság National Health Service (NHS) Foundation Trust egészségügyi szolgáltatója 2015-ben egyezményt kötött a DeepMind Health céggel, melyben díjmentes hozzáférést biztosított a cégnek mintegy 1,6 millió, személyesen azonosítható egészségügyi adatrekordhoz. Később azt állították, hogy az NHS nemzeti egészségügyi szolgáltatót „elkábította az algoritmusgyártó cég csodája, és a jövőben legalábbis igyekeznek majd jobban kontrollálni az adatokat és az átláthatóságot. Amit az NHS nem vett észre: ők voltak a birtokosai a ténylegesen értékes és fontos tényezőnek, azaz az adatkészletnek” [36].

A nyílt, szabadon hozzáférhető adatok hatalmas előnyöket kínálnak a páciensek és a közjó, a társadalom és a gazdaság számára egyaránt. Naivitás lenne azonban azt várni, hogy az adatok tulajdonosai díjmentesen közreadják értékes erőforrásaikat. A francia French Radiological Society (SFR) szövetség 2018 évi közgyűlésén egy „DRIM France AI” elnevezésű, AI alapú ökoszisztéma megalapozását jelentették be. A terv elképzelése, hogy 5 éven belül több mint 100 millió orvosi képet tartalmazó, minősített adatbázist hozzanak létre. Ezt az adatbázist használhatják AI eszközöket fejleszteni szándékozó cégek, mely AI eszközöket díjmentesen használhatják majd francia kórházak és radiológusok. Legalább abban lenne célszerű az országnak egyetértésre jutni: milyen típusú adatmegosztás legitimált és szabályozható, és miként valósulhat meg az adat-előállítók, adat-tulajdonosok, menedzserek és felhasználók biztonságos és igazságos adatmegosztása.

A kibocsátásra/hozzáférésre vonatkozó információk és az adathasználati egyezmények (DUA; Data Use Agreement) kritikus fontosságú eszközök az adatok alkalmazásának transzparenciája és etikussága tekintetében. A DUA egyezmények rögzítik, hogy a szerződő felek közül ki mit tehet és mit nem tehet az adatokkal, és arról is rendelkeznek, mit történjen az adatokkal az egyezmény időtartamának lejártá után. A DUA egyezményeket rendszeres időközönként frissíteni, aktualizálni szükséges, hogy a páciensadatok esetleg új típusú használatát megfelelően szabályozzák. Maguk az adatok önálló entitásoknak tekinthetők, értéküket flexibilitásuk nagyban befolyásolja. Minél többcélúan alkalmazhatók, sokoldalúbban megoszthatók és kombinálhatók az adatok, annál értékesebbek. Ahogy az adat felhasználásának módja, és velük az adat is változik, annak megfelelően minden egyes adatállapotot dokumentálni kell. A DUA egyezmények bizonyos esetekben korlátozásokat írhatnak elő az adatok újrahazsnosításával kapcsolatban, hogy ezzel az adatok bizalmassága elleni vétségek megelőzhetőek legyenek. Ezzel megakadályozható az is, hogy az adatok alkotta adatbázisnak AI algoritmus tréningre történő alkalmazása során „adateltérés”, adattorzulás lépjen fel. A jövőbeli DUA egyezményeknek rendelkezni kell az adatverziók kezeléséről is - különösen, ha az adatokat AI modellek tesztelésére, trenírozására vagy validálására alkalmazzák. Javasolt, hogy az egyezmény tartalmazzon új és átfogó szabályokat az adatok újrafelhasználásáról és a szellemi tulajdonról is. Az egyes adatokat fogadó intézmények felelősséget kell vállaljanak a kapott adatok eredetének ellenőrzése

kérdésében, és teljes mértékben meg kell értsék az adatokhoz kapcsolódó szabályokat és engedélyeket. Javaslatok vannak arra, hogy vizsgálatok előtt minden páciens írjon alá DUA egyezményt bármilyen harmadik féllel (ha van ilyen), aki/amely hozzájárul a páciens digitális adatainak létrehozásához. Mindezzel biztosítható és nyomon követhető az adatok minősége, biztonsága és használata [37]. Az adatkezelésben egy másik megközelítés a dinamikus „beleegyező hozzájárulás”, ez egy elektronikus eljárás, mely a kutatók és a kutatásban résztvevő alanyok között folyamatos kommunikációt tesz lehetővé [38].

Hangsúlyozzuk véleményünket: a kizárólagos adathasználati jogokat tartalmazó DUA egyezmények etikátlanak, mivel ezzel a megállapodás jelentős mennyiségű radiológiai adatot zárhat ki az általános, a köz javát szolgáló felhasználásból. Az ilyen egyezmények a hatalom koncentrációját növelik és a transzparencia szempontjából hátrányosak. A kizárólagos adathozzáféréstől szóló szerződések nem a közjót szolgálják.

Az intézményi felügyeleti testületek (IRB; Institutional Review Board, más néven független kutatás-etikai bizottság) követelményeinél is szükséges, hogy tartalmazzanak a páciensadatok új típusú használatára vonatkozó szabályokat. Egyes, főként az USA-n kívüli IRB-k szabályozása lemond a beleegyező nyilatkozat követelményéről, ha az ilyen nyilatkozat nem megvalósítható, vagy akadályozná valamely kutatási tanulmány vagy AI modell validációját. Vajon milyen az a helyzet, melyben nem élvez abszolút elsőbbséget a páciensadatok bizalmas jellege, melyben a páciens érdekeit felülírhatja más érdek, és mikor alacsony egy ilyen adathasználat kockázata, esetleg amikor a kényszerítő közérdek erősebb lehet [39]? Márpedig ha ilyesmi jelentkezne, a páciensekben ezt tudatosítani kell.

A nagymennyiségű orvosi adat megosztására és kezelésére szolgáló műszaki infrastruktúra iránti igény ösztönzi az új, robusztus támogató technika kialakulását. Elméletileg az ún. blokklánc (blockchain) elosztott adatbázis technológia eléggé hatékony és átfogó módszereket kínál különböző helyeken lévő egyének és intézmények számára az orvosi adatok számítógépes hálózaton történő biztonságos koncentrációjához, azok egyszerű hozzáféréséhez [40, 41]. Ennek a technológiának részletei és problémái jelen közlemény hatókörén kívül esnek.

A teljes átláthatóság és maximális bizalom érdekében előnyös lenne a páciensadatok jelentőségét tudatosító keretrendszer létrehozása, útmutatással a különféle adathasználati módokhoz. Mit kell tenni a radiológiai szakma a páciensek bizalmának elnyeréséhez, ahhoz, hogy meggyőzze a pácienseket: adataik felhasználása megfelelően etikus módon történik? Miként segíthet a radiológia a pácienseknek akkor, ha azoknak adataik pénzügyi vagy egyéb értékével kapcsolatban követelések támadnak? A pénzügyi értékkel kapcsolatos követelések jórészt jogi precedencia alapon kezelhetők, és kevésbé vizsgálatok etikai szempontjából. Mindez ráadásul országonként eltérő módon szabályozott. A felmérések szerint a legtöbb páciens hajlandó beleegyezését adni adatainak megosztásához [42], és valószínűleg bízik benne, hogy adatait megfelelő módon kezelik és használják. Azonban az adatok megosztásának indítéka - hogy akadémia közegben kutatás, vagy iparban kereskedelmi termékfejlesztés a cél - erősen megváltoztatja az alanyok véleményét és szándékát az adatok szabad felhasználásával kapcsolatban [43]. Lehet, hogy a helyzet jövőben változik majd: ha például a kutatásban a páciensadatok bizalmassága sérül, vagy ha a páciensekben tudatosan adataik akár pénzügyben is kifejezhető

értéke [44]. Egyre nagyobb mértékben kerülnek személyes páciensadatok begyűjtésre nem orvosi helyzetekben. A páciensek, vagy akár az egész közösség felkérését kaphatnak radiológiai vizsgálati eredményeik megosztására, vagy akár áruba bocsátására [30, 45]. Jelenleg még nem létezik egységes állásfoglalás a beleegyező nyilatkozatokra vagy adatkezelési szerződésekre, melyek az ilyen adatok használatát és újra felhasználását szabályoznák. Ugyanígy nincsenek követelmények arra, miként kellene a pácienseket értesíteni: hogyan és kik használják adataikat?

A társadalom tagjainak a radiológián kívül is rengeteg könnyen személyhez köthető adata van. Ezek között szerepelnek egyéb orvosi adatok, leletek, patológiai vagy genetikai laboreredmények, mobiltelefon használatból származó adatok, vagy akár személyes testedzési adatok a fiziológiai állapotot és mozgást követő sok elektronikus eszköz révén. További példák az adatok sokaságára az internetes keresésekből, használati és vásárlási szokásokból nyerhető információk, a videokamerás megfigyelésből, útvonalkövetésből származó adatok vagy akár környezeti, például időjárás feljegyzések. Ezek közül az adatok közül sok nyilvánosan elérhető, és elméletileg fennáll a lehetőség, hogy az információ összesítésével széleskörű, mondhatni „360 fokos” körképet kapjunk az egyénről, azaz a páciensről. Az ilyen integrált adathalmaz esetleg pontosabb diagnózist és kezelést tesz lehetővé az egyén számára, azonban az ilyen forrású adatok anonimizálása szinte lehetetlen, feldolgozásuk a bizalmasság megsértésének jelentős kockázatával jár.

A páciensek a legritkább esetben tudják, mi is történik adataikkal. A bizalom megnyerésének fontos útja az átláthatóság biztosítása. Ha az egyén teljes körű információt kap egy intézmény adatgyűjtési és kezelési gyakorlatáról, részt vehet az adatok gyűjtésében, vagy akár irányíthatja is az adatgyűjtést, bizalma megerősödik az intézmény iránt, esetleg egyfajta „márkahűség” is kialakul a megbízható, átlátható szervezet irányában. Ugyanakkor maga az intézmény is tisztában kell legyen az adatmegosztás és adat újrafelhasználás céljaival, ami alapvetően fontos minden etikus információkezelésnél, és kiemelten lényeges AI fejlesztéseknél. Sok minden függ a kontextustól: ha a páciens saját adatait olyan környezetben és összefüggésben találja meg, ahol erre nem számít, akkor az egyén meglepetése könnyen bizalmatlanságra válthat.

## Adatok bizalmassága

Az egyének privát szférához, bizalmassághoz való joga az „egyedül maradáshoz” való jogként értelmezhető, ami egyenértékűnek tekinthető azzal, hogy sem más személyek, sem szervezetek részéről nem áll megfigyelés alatt [46]. Ebben a felállásban csak jogosított személyeknek lehet hozzáférése a páciensadatokhoz. Minden jogos és méltányos erőfeszítést meg kell tenni annak érdekében, hogy ez a bizalmasság megvalósuljon és fennmaradjon. Különösen fontos ez akkor, ha az adatokat újra felhasználják, azokat különféle adat-tulajdonlási és felelősségi láncolaton keresztül továbbítják.

Az Amerikai Egyesült Államokban a Health Insurance Portability and Accountability Act (HIPAA; az egészségbiztosítás transzferálhatóságával és elszámoltathatóságával kapcsolatos szabályozás) szigorú bizalmas adatkezelés ír elő a védett páciensazonosító információk (PHI; Protected Health Information) számára. Ezért a páciensadatokat gyakran

névtelenítik, azaz az egészségügyi információk melletti személyi azonosítókat kitakarják vagy eltávolítják, mielőtt az adatokat kutatási vagy kereskedelmi célokra használnák [47]. Az orvosi képek speciális anonimizálási problémákat vetnek fel. Például a fejről és nyakról készült részletes rétegfelvételek összesítésével 3D modell alkotható, melyet egy arcfelismerő szoftver képes lehet azonosítani [48].

Név nélküli röntgenfelvételek tartalmazhatnak nyakláncokon vagy karkötőkön lévő azonosítási információt, vagy a képbe kerülő implantátum eszközök (pl. defibrillátor vagy pacemaker) sorozatszámja is azonosíthatja a személyt [49]. Az ultrahang képek tartalmazhatnak „beégetett”, azaz pixeladatok közé nehezen eltávolítható módon beiktatott azonosítókat. A radiológiai felvételek kiterjedt meta-adatokat is tartalmaznak, melyek némelyike azonosíthatja a páciens. A DICOM szabványú képeknél az ún. „private tag” egyedi, vizsgálóberendezés gyártótól függő jelölőmezői (melyeket gyakran nem dokumentálnak kellően) váratlanul tartalmazhatnak az intézményt vagy páciens, esetleg annak betegségét azonosító információt.

Ha ezeket az adatokat specifikus tartalmak kinyerésére, vagy AI algoritmusok tanítására használják, akkor a modell ezeken az adatokon „edződik”, és esetleg nem lesz képes általánosításra, ha ezek a specifikus adatok nem állnak rendelkezésre egy másik szituációban. Ma már az a helyzet, hogy a radiológiai vizsgálati eredmények teljes és valódi anonimizálása a DICOM képek név adatmezőinek, tag-jeinek törlésénél vagy átírásánál többet igényel, esetleg a képek humán megfigyelő általi áttekintésére is szükség lehet. Az Amerikai Egyesült Államokban több kutatóközpont eljárásrendjében előírja: páciensadatok nem oszthatók meg addig, amíg kettő független emberi szemlélő egyenként meg nem tekintette a képadatokat, és el nem távolította azokból az azonosítókat.

Persze a radiológiai vizsgálati adatok és egyéb orvosi információk névtelenítésének itt leírt hagyományos, szigorú gyakorlata sem abszolút érvényű módszer. A korábban leírt „360 fokos” teljes körű megközelítés alkalmazásával a nagy tömegű adat kezelésében jártas és ügyes intézmények nagy valószínűséggel majdnem minden anonimizált radiológiai vizsgálatot ismét azonosítani tudnak [50]. Technikailag lehetséges, hogy egy nagy szociális média cég a személyek okos-telefon használatából és egyéb személyes eszközökből adatokat gyűjt, majd ezeket az online keresési vagy vásárlási műveletek naplózásával együtt összeilleszti az egészségügyi adatokkal. Ennek révén megfelelően célzott reklámokat küldhetnek az adott személynek, vagy esetleg eladhatják ezeket a kinyert és összesített adatokat bárkinek, mondjuk biztosítótársaságoknak vagy ápolóotthonoknak. A radiológiai szolgáltatók számára hasznosak lehetnek az ilyen adatkészletek, például annak megítélésére: kinek milyen jövőbeli vizsgálatra lesz szüksége. Az ilyen jellegű, mindenre kiterjedő információösszesítés és adathozzáférés tovább növeli az adatbiztonság és adatvédelem jelentőségét és szükségességét. Mindig fennáll a veszély, hogy az „adatpiac” egészségügyi adatokhoz is hozzáférő, rosszindulatú szereplői zsarolásnak teszik ki azokat a személyeket, akik orvosi adatai között szerepelnek olyanok, melyeket nem kívánnak nyilvánosságra hozni.

Az etikus módon eljáró egészségügyi szolgáltatók mindent megtesznek az adatok bizalmosságának és biztonságának megőrzése érdekében, de eközben persze átláthatók is kell legyenek. Ennek során felmerülhet: az orvosi adatok esetleg nem maradhatnak mindig abszolút privát jellegűek. A tökéletes anonimizálás mindenestre egyelőre a legjobb



megoldás.

AI algoritmusok trenírozására történő adathasználat az adatok kitétségének egy újabb aspektusát jelenti. Az elterjedten használt ún. „deep learning” (többrétegű, adaptív gépi tanulás) eljárások gyakran tartalmazznak a trenírozáshoz használt adatok részleteit is feldolgozó módszereket. Az algoritmus viselkedése esetenként váratlanul és szándékolatlanul felszínre hozhatja ezeket a részadatokat [51]. Szélsőséges esetben, akár rossz szándékkal tervezhető olyan algoritmusok, melyek szándékosan keresnek és szivároztatnak ki érzékeny adatokat. Ezt az eljárást - a számítógépes vírusok szakzsargonjából vett „intentional back-dooring”, azaz „szándékos hátsó-ajtó kinyitó” technikaként ismeretes [52]. Ezért minden AI alkalmazást ugyanúgy kell kezelni, mint minden más szoftver megoldást, azaz velük kapcsolatban is a számítógépes programokra alkalmazott intézményi biztonsági eljárásokat kell alkalmazni.

## Adatmódosulás és torzulás

Az adatmódosulás vagy adattorzulás (bias) szisztematikus eltérés az igazságtól. Ilyen eltérés akkor jön létre, ha egy adatminta elemei nem semlegesen reprezentálják a valóságot. Tipikus esetek a radiológiai AI alkalmazásoknál a leletezés, a kiválasztás és az automatizálás során létrejövő adattorzulások. Az automatizálási adattorzulás kérdését „A gyakorlat etikája” részben tárgyaljuk.

Leletezési adattorzulás akkor jelentkezik, ha a leletezett és annak során értelmezett adatok nem teljesen és nem kellően semlegesen reprezentálják a tényleges állapotot, mert egyes adatkomponenseket szelektív módon kizárnak. Ez a klinikai adatok gyakorlatában például abból adódhat, hogy a pozitív kutatási eredmények általában jobban hozzáférhetőek (azaz felülreprezentáltak), vagy, mert bizonyos esetek duplikáltak, esetleg többszörösen „túl-leletezettek”. Ezzel szemben a negatív kimenetelű tanulmányok adatai jellemzően alulreprezentáltak, a leletezésben nem kellő gyakorisággal szerepelnek. Ilyen torzulási jelenség a prototípus adatok előfeltételezése, példa erre banánok leírása esetén az, amikor a sárga jellemző színt nem említik attribútumként, mert egyéb megjelölés hiányában az eleve feltételezett [53].

Kiválasztási vagy mintavételezési adattorzulás jelentkezik, ha a begyűjtött adatok nem kellő pontossággal reprezentálják a lakosságot [54]. Ennek gyakran oka az, hogy csak a könnyen hozzáférhető, rendelkezésre álló, vagy érdekes adatokat vonják be az adatkészletbe. Például egy adott intézményből szerzett, AI rendszer trenírozására használt radiológiai adatok pontosan reprezentálják az intézmény hatókörének pácienseit, de nem eléggé reprezentatívak a szélesebb lakosság vonatkozásában, melyre az AI modellt használni szándékozzák. Lehet, hogy az ilyen adatkészlet akarva-akaratlanul diszkriminálja a mintavételben nem kellően jelenlévő részpopulációt [55].

Kiválasztási adattorzulás felléphet nyíltan (ismert módon) és szándékolatlanul. Például, ha egy specifikus betegség felismeréséhez tervezett AI algoritmus trenírozásához használt adatbázis összes képe olyan páciens-csoporttól származik, melynek tulajdonságai eltérnek a teljes lakosság tulajdonságaitól, akkor a teljes lakosság feldolgozására szánt AI modell

szisztematikusan helytelen válaszokat fog adni olyan alanyok esetében, akik tulajdonságai nem egyeznek a trenírozáshoz használt adatbázis alanyainak tulajdonságaival. A megválaszolendő kérdéstől függően, az adatkiválasztáshoz alkalmazott releváns tulajdonságok lehetnek fizikai adatok vagy orvosi jellegű információk, mint pl. életkor, nem, testmagasság, genetikai és orvosi előtörténet, de lehetnek gazdasági, szociális helyzetet, vagy a képzettséget leírók is. Mivel az AI algoritmusok jellemzően az emberi értékelőnél nagyobb adathalmazból dolgoznak, és a tulajdonságokat „durvább” felbontásban nyerik ki az adatokból, ezért gyakran nehéz előre meghatározni, hogy a trenírozó adatbázis mely tulajdonságai okoznak majd adattorzulást, vagy vezetnek egyébként nem etikus AI modell kialakulásához.

Az adatkészlet-eltolódás (DS; dataset shift) jelenségek a kiválasztási torzulás egy alcsoportját képezik. Ez a hiba ma az AI széleskörű alkalmazásának útjában álló legfontosabb akadályok egyike. DS jelentkezik a legtöbb a radiológia szituációban, mivel az AI modell trenírozásához használt képadatok nem pontosan reprodukálják a későbbiekben, klinikai környezetben fennálló feltételeket. Ezek közé tartozik az AI modell kísérleti fázisában az algoritmusba jutó adattorzulás, melynek oka például szintetikus vagy augmentált trenírozó adatok használata lehet. Más szavakkal: az ilyen adatokkal végzett előzetes trenírozás nem elegendő, vagy nem megfelelő a modell számára ahhoz, hogy új helyzetekben is pontos előrejelzést szolgáltatson [56].

Míg a gyakorlott radiológus könnyűszerrel észreveszi, és a leletezésben alkalmazkodik az olyan képekben meglévő különbségekhez, mint a szeletvastagság, a szkanner típusa vagy gyártója, MRI esetében térerősség, a gradiensek tulajdonságai vagy a kontraszt időzítésének eltérése, addig az AI általában nélkülözi ezeket a képességeket. Ha például egy AI algoritmust 3T MRI képeken treníroznak, lehetséges, hogy az 1,5T MRI vizsgálatból származó képekre nem ugyanazokat az eredményeket hozza. Ugyanez igaz a fentebb felsorolt egyéb képalkotási paraméterekre is. A DS adat-eltolódás jelenség enyhítésére az egyik megközelítés a széleskörű tréning, szigorú validálás és megfelelő teszt körülmények alkalmazása. Ilyen megközelítést alkalmaz a szepsziseket detektálni hivatott InSigh rendszer [57, 58]. Egy másik módszer az adatokat felismerő, normalizáló és transzformáló matematikai eljárásokat alkalmaz a DS adat-eltolódás minimalizálása érdekében.

Az adatkészletek statisztikai jellegű torzulása eléggé általánosan fellépő jelenség ahhoz, hogy az AI algoritmusok különbséget tudjanak tenni a különféle adatbázisok között. A kézi módszerrel kiválasztott adatok alapvetően több adattorzulást eredményeznek, mint az automatikusan vagy véletlenszerűen kiválasztottak. Mintavétel vezérlés közben ún. felügyeleti adattorzulás léphet fel, mikor például a vizsgálatvégző megválaszthatja a képalkotás nézeti szögét - ami gyakori ultrahang vizsgálatnál. A negatív adatkészlet-torzulás akkor jelentkezik, ha az adathalmazban egyes pozitív vagy egyébként érdekes vizsgálatok túltreprezentáltak, azaz tényleges előfordulási gyakoriságuknál nagyobb súllyal szerepelnek a minták között. Ez a kérdés különösen komplex a radiológiában, ahol a vizsgálatok túlnyomó többsége normál, azaz negatív, elváltozást nem tartalmazó. Ez esetben szükséges lehet ezt kiegyensúlyozni azáltal, hogy megfelelő számú patológiás mintát is beválogatunk az adatkészletbe: Mindezt anélkül kell megtegyük, hogy abnormális módon felborítanánk az egyensúlyt, és adattorzulást okoznánk az algoritmus számára. Ha elegendő mintát kívánunk egy ritka patológia prezentálásához előállítani, és ehhez szintetikus vagy augmentált adatokat használunk, akkor fennáll annak veszélye, hogy ezek a mesterséges befolyásoló tényezők helytelen módon

adattorzulást okoznak az adatkészletben.

A radiológiai adatkészletek gyakran kiegyensúlyozatlanok, ami azt jelenti: egyes adatkategóriákból sokat tartalmaznak (főként normál vizsgálatok eredményei), és csak kisebb arányban szerepel bennük mint az adott patológiával. A kiegyensúlyozatlan adatbázisokban egyes kategóriák alul-mintavételezettek, mások pedig felül-mintavételezettek lehetnek. Mindez előállhat azért, mert törekszünk az AI modell hatékonyságát növelni, a programfuttatás idejét csökkenteni, azonban ez az egyensúlyzavar adattorzulást eredményezhet.

Néha úgy vélekedünk, hogy az adateltérés vagy adattorzulás etikailag semleges kategória. Ilyen megvilágításban az adateltérés akár jótékony hatású is lehetne. Ha egy egészségügyi rendszer az elvártnál aránytalanul alacsonyabb színvonalú szolgáltatást nyújt egy adott népességcsoport számára, lehetőség van az egyenlőtlenség ellen hatni AI eszközökkel, amely a jó egészségügyi eredménynek prioritást ad az adott népességcsoporthoz tartozó minden páciens számára. Ennek ellenére véleményünk az, hogy az adateltérés, adattorzulás jelenség negatív dolog, és a radiológia helyes etikai megközelítése az adattorzulások minimalizálása kell legyen.

## Adatok jelölése és alapigazság

A klinikai radiológiában jelenleg alkalmazott AI modellek úgynevezett felügyelt gépi tanuló algoritmusok (sML; supervised Machine Learning). Ennek során a modell olyan jól megtanulja jelölők adott képekhez történő hozzárendelését, hogy ha a modellnek új képeket mutatnak be, akkor jó eredménnyel megjósolja: melyik új képhez melyik jelölőt kell rendelni. Ez különösen fontos akkor, ha a jelölők ún. alapigazsághoz rendelték – esetünkben ez az alapigazság állapotinformáció, a páciensnél megtalált patológia vagy annak hiányára vonatkozó kijelentés.

Az alapigazság meghatározása gyakran nem könnyű a radiológiai képalkotásban. Például egy AI algoritmus trenírozható arra, hogy a csukló sajkacsontját ábrázoló radiológiai képeken törést vegyen észre. Az AI trenírozására használt alapigazság jelölő származhat egy radiológus észleléséből, mely a csontot törésesnek vagy épnek leletezi. Egyes törések olyan parányiak, hogy az eredeti felvételen nem észlelhetők, vagy észlelhetők, de a radiológus nem veszi észre őket. Az apró, de a csuklóban jelentős szerepű sajkacsont törésére irányuló klinikai radiológiai kérdés esetén, ha az első felvétel negatív, de a páciens két hét múlva is még fájdalmáról panaszkodik, akkor megismétlik a felvételt. Lehet, hogy a törés az első vizsgálatnál rejtve maradt, de ezen a második felvételen rendszerint jobban észlelhető. Vajon az első röntgenfelvétel észlelését vegyük alapigazságnak, vagy helyesebb, ha az alapigazság paradigma tartalmazza a második röntgenfelvétel előírását azzal, hogy kérdés: mit mutat ez a második (esetünkben szenzitívebb) vizsgálat? Más szavakkal: melyik klinikai eredmény a legfontosabb? Egyes radiológiai vizsgálatoknál az alapigazság jelölőt nem a leletből vett információ adja, hanem inkább egymást követő képalkotó, fizikai és klinikai észlelések, sebészeti beavatkozás eredmények, patológiai és genetikai laboradatok sorozata.

Minden radiológus hibázhat: nemcsak hogy nem leletez 100%-ban helyesen minden vizsgálatot, hanem előfordulhat, hogy másnap már eltérően leletezi a képeket, vagy a más radiológustól érkező felvételeket. Az egyetlen szakértőtől származó kvalitatív besorolás alapján hozzárendelt alapigazság összezavaró lehet az egy megfigyelőn belüli, vagy megfigyelők közötti variabilitás miatt (intra- vs. inter-observer variability). Több radiológus interpretációjának a meghatározásba történő bevonása javítja a jelölők pontosságát [59]. Ha három radiológus leletezne minden vizsgálatot, a többségi vagy konszenzusos eredményből lehetne alapigazságot kialakítani – persze a gyakorlatban ez kivitelezhetetlenül költséges lenne.

Alternatív lehetőségként szemi-kvantitatív besorolási rendszer alakítható ki a képadatok alapigazságának meghatározásához, amelyhez osztályozási atlaszokba rendezett szigorú szabályok tartoznak, és amelyeket több leletező észlelései és elemzése alapján hoztak létre. Az ilyen kép-alapú besorolási rendszerek értékelésére formális technikákat hoztak létre, egy ilyen például az OMERACT Filter [60] szűrőeljárás. Egy AI rendszer akkor értékelhető sikeresnek, ha legalább olyan jók az eredményei az ilyen besorolási feladatok végzésénél, mint a humán értékelő-leletező szakembernek. A fentebb említett sajkacsont-törés példánál maradva, elképzelhető egy szemi-kvantitatív besorolási rendszer, amely különféle jelenségek alapján - mint például csontfelszín szakadás, fényes vonal jelenléte a képben, csontdenzitás változásai, csuklócsontok illeszkedése - rendel pontszámokat a döntéshez.

A fenti példák illusztrálják az AI algoritmusok trenírozásához használandó, alapigazság jelölőkkel ellátott adatbázis létrehozásának nehézségeit. Milyen szempontokon alapuljon a döntés, és ki határozza meg ezeket a szempontokat? Az adatbázisokba mélyen beépülő adattorzulások elkerülése érdekében biztos, hogy az erre adott válasz függ majd a specifikus radiológiai döntési feladattól. Ugyanígy biztos, hogy rendkívül körültekintően kell megfontolni létrehozását, és főként előre - *a priori* - meg kell határozni az elveket.

Az etikus megközelítés azt sugallja a megoldást mérlegelő számára, hogy a lehető legjobb alapigazság jelölők létrehozása mindenképp prioritást kell élvezzen az eljárás fejlesztésének költségeivel és nehézségeivel szemben. Továbbá átláthatóságot kell biztosítani minden egyes adatbázis alapigazság-meghatározással kapcsolatban. Az adatkészletek esetleges statisztikai torzulásainak észlelése és jelentése, és az ezzel kapcsolatos standardok bevezetése jó szolgálatot tenne a radiológia és az orvostudomány számára. A radiológiai közösség igenis kérdéseket kell feltegyen a rendelkezésre álló adatok tárgyában, transzparenciát kell biztosítson az értékelési folyamat és a kérdésekre adott válaszok ügyében. Ez különösen fontos akkor, ha nyilvánosan hozzáférhető adatbázisokat alkalmaznak AI tréninghez, mivel a kutatók és alkalmazók esetleg nincsenek tudatában az adatok mögötti feltételezéseknek vagy nincsenek ismereteik az adatokban rejlő torzulásokról.

Egy AI modell bevezetésénél a felelősöknek az alábbi és hasonló kérdésekre kell válasz találniuk:

- Milyen (statisztikai) adattorzulás lehetséges az adatkészletben?
- Mit tettünk az adatkészletben esetleg rejlő torzulás felderítése érdekében, és miként befolyásolja az a modellt?
- Mik az adattorzulás hatásának lehetséges kockázata, és mit tettünk a kockázatok

enyhítése érdekében?

- Milyen reziduális adattorzulás maradhat a rendszerben, és a későbbi felhasználók hogyan vegyék figyelembe ezeket a maradék adateltéréseket?
- Vajon megfelelő-e az alkalmazott alapigazság-jelölő módszer az adott klinikai kérdések esetére?

## Algoritmusok és tanuló modellek etikája

Hatásmechanizmusának legmélyén minden AI modell besorolási, osztályozási rendszereket alkalmaz az eredmény eléréséhez. Az AI döntéseinek első és talán legegyszerűbb megközelítése - példaként - a következő formális logikai modell: „Ha egy egyébként egészséges betegnek láza van, akkor lehet, hogy fertőzése van”. Egy másik szemlélet inkább valószínűségi természetű, Bayes-i eszköztárt alkalmaz: „Ha egy páciensnek láza van, akkor fertőzésének valószínűsége legyen X%”. Egy harmadik eljárás a hasonlóságok általánosítása alapján alkot előrejelzést: „Az adott pácienshez leginkább hasonló testhőmérsékletű, szimptomájú, életkorú és egyéb adatú egyedek alkotta csoport elemzése révén arra a következtetésre lehet jutni, hogy azoknak az egyedeknek X%-a szenvedett fertőzésben”. Van egy negyedik, a neuronok működését tükröző modell, a neurális hálózat. Ilyen például a „deep learning” módszer, melyben az adatokon történő trenírozás a modell alkotta „neuronok” kapcsolati erősségét megváltoztatja.

## Gépi döntéshozatal

A döntéshozatal tulajdonképpen választás egyes feltételezések vagy cselekvési alternatívák között. A döntés rendszerint cselekvést vált ki. A humán döntéshozatali folyamat az adott személy ismeretein, értékszempléletén, preferenciáin és meggyőződésén alapuló, alternatívák közötti választási folyamat. Az AI algoritmusok döntéseiknél, a rendelkezésre álló bemeneti adatok tulajdonságai alapján, alternatívák közül választanak. A felügyelt gépi tanulásnál (supervised Machine Learning) az AI algoritmus - alternatívák közötti választása során - előzetes trenírozás alapján adattulajdonságokat és jelölőket párosít össze. Az ilyen jelölők alkalmazása ott gyakori, ahol a humán értékszemplélet, emberi preferencia és meggyőződés könnyen átvihető a gépi közegbe. Ennek során gyakran jelennek meg a humán értékítéletből transzferált adattorzulások is.

Míg az AI jól teljesít a besorolás és osztályozás típusú feladatoknál, nehezen birkózik meg az olyan absztrakt koncepciókkal, mint például a pártatlanság és korrektség, vagy az egyenlő bánásmód [13]. Ehhez járul még, hogy a mesterséges intelligencia hozzáférhetősége és alkalmazása nem az AI természetes velejárója. Ezeknek a problémáknak a felelőssége az emberre hárul, akinek előre kell látnia a gyorsan fejlődő AI modellek esetleges jövőbeli tévműködését vagy téves használatát, de a velük kapcsolatos potenciális visszaéléseket is. Védelemmel kell szolgálnia a lehetséges káros következményekkel szemben, ideális esetben még mielőtt még azok bekövetkeznének [61].

Az AI modell algoritmusokból, és azokat „trenírozó” adatokból áll. Az algoritmusok fejlesztése

során azok követhetősége és tesztelhetősége megköveteli, hogy a feldolgozott adatkészletek pontos verziója ismert, tárolt és reprodukálható legyen. Elméletileg konstruálható olyan AI modell, mely új adatokon történő tanulás révén folyamatosan változik. A jelenlegi AI modellek trenírozása gondosan összeállított adatkészletek segítségével történik, majd egy adott „tanultsági” állapotot mintegy befagyasztva, az algoritmus kimerevített változata kerül implementálásra. Ha az AI modell nagy kockázatú döntések ügyében felelős, nem valószínű, hogy a folyamatos továbbtanulás révén nyerhető marginális előny nagyobb, mint az esetleg szándékozatlanul bekövetkező teljesítményromlás kockázata. Az ilyen befagyasztott és dokumentált AI modell munkaverziók alkalmazásának módszere (version control, változatok követése) a jelenleg standard gyakorlat. Sajnos ugyanezt a következetes szigorúságot eddig nem alkalmazták a trenírozó adatkészletekre. Hasonló szigorú minőségbiztosítási és etikus szoftver-kezelési eljárásrendet lenne javasolt mostantól bevezetni az adatokra is. Ez kritikus probléma, mivel AI rendszerek teljesítőképességi problémái esetén szinte lehetetlen az eredeti okot kideríteni, és korrekciót szolgáltatni anélkül, hogy pontosan ismernénk az alkalmazott adatbázisokat.

## Algoritmus-választás

Minden AI megoldás fejlesztésének kezdeti lépései a trenírozó adatok megfelelő megértése, a modell kérdésselvetésének meghatározása, és a lehetséges adattorzulás előzetes, kritikus értékelése. Az algoritmus kiválasztása függ az adatok természetétől, a végrehajtásra szánt processzálsági időtől, és a végrehajtandó feladattól. Egyes algoritmusok hatékonyabban működnek kis adatkészleteken, míg mások inkább nagyobb mintasereget igényelnek. Képfelismerési célokból ún. konvolúciós neurális hálózatok (CNN: Convolutional Neural Network) mutatják a legbiztosabb eredményeket. A fejlesztők a kérdésselvetés és/vagy a tréningadatok tulajdonságai alapján választják ki az algoritmus struktúrát (például lineáris vagy nemlineáris). Annak megértésén és eldöntésén túl, hogy milyen algoritmus a legmegfelelőbb az adott kérdéshez, további etikai problémák jelentkeznek: milyen algoritmus adhatja a leghasználhatóbb eredményt a páciensek kezeléséhez, ehhez mérten mik a számítási apparátus költségei és korlátai, esetleg milyen korlátozások lehetnek a rendelkezésre álló trenírozó adatok mennyisége és milyensége tekintetében.

A modell célkitűzése is bevihet torzulást. A tréning adatkészlet kiválasztásánál a radiológusnak mérlegelni kell az esetleges akaratlan következményeket, és előre értékelnie kell a modell teljesítőképességét, annak korrektségét különféle pácienscsoportok vonatkozásában. Ezt azzal lehet legjobban biztosítani, ha a modell által a gyakorlatban majd feldolgozandó adatok jól illeszkednek a modell teljesítőképességének validálásánál alkalmazott tanulási és tesztelési adatokhoz. Ez a folyamat hasonlít a bizonyíték alapú orvoslásnál alkalmazott elvekhez, ahol a diagnosztikai eredmények alapján választ az orvostudomány terápiás módszert.

Az algoritmusok egyre növekvő komplexitása, és az AI modellek fejlesztéséhez és trenírozásához szükséges megfelelő személyzet hiánya miatt új fejlődési irány jelentkezik: az automatizált gépi tanulás (Automated Machine Learning, AutoML). Az AutoML lehetővé teszi a számítástechnikában kevésbé jártas, praktizáló radiológusok számára is AI modellek létrehozását és trenírozását. Bár ebben a folyamatban nagy lehetőségek vannak az AI

„demokratizálása” tekintetében, a nem kellően gyakorlott trenírozók - az AI modellek „fekete doboz” jellege miatt - nem mindig eléggé tájékozottak a feladat összetettségének és buktatóinak kérdéseiben. Ahogy a radiológusok egyre fokozottabban lesznek felelősek az AI eljárások fejlesztése és felügyelete vonatkozásában, ezzel arányosan kell megfelelő ismereteket szerezzenek az AI lehetséges etikátlan, adattorzulásos eseteiről vagy egyéb tévműködéséről.

## Algoritmus-tanítás

Miután az algoritmus trenírozást kapott egy adatkészleten, ML gépi tanulási modell lesz belőle. Ez a lépés is bevihet adattorzulást, mivel az algoritmus megörökli az adatok kiválasztásánál és előkészítésénél alkalmazott döntések következményeit. A lehetséges adattorzulás (különösen a DS adateltolódás) minimalizálása, és a páciensek érdekeinek optimális figyelembevétele érdekében kritikusan fontos, hogy a modell trenírozásához alkalmazott adatkészlet valóban jól reprezentálja a klinikai döntési gyakorlatban majd eléje kerülő eseteket. Ez gyakran intézmények közötti, megosztott adatbázisokon történő trenírozást tesz szükségesé. Az adatmegosztás problematikájának kihívásait ismerve, többféle akadály is korlátozhatja az AI tréninget. Ha a megosztott adatkészleten történő trenírozás jogi és bizalmas adatkezelési akadályai jelentősek, a fejlesztők választhatják az FDA (Amerikai Egyesült Államok orvostechikai felügyeleti szerve) engedélyezéshez meghatározott minimális AI tanulási modellt. Hasznos megközelítés az adatkészletek megosztása helyett a modellek súlyozásának és paramétereinek kommunikációja intézmények között, mivel ez utóbbiakra nem vonatkoznak bizalmas páciens adatkezelési előírások.

## Modellértékelés és tesztelés

Ha az AI modell trenírozása befejeződött, különféle adatokkal tesztelik, hogy megállapítsák: milyen jól működik. Ennek során megvizsgálják, hogyan dolgozza fel az atipikus bemeneti adatokat, vagy az olyan adatokat, melyeket feltételezhetően nem képes jól kezelni. A modell-teszteléshez tartozik a megfelelő tesztelő adatok kiválasztása, a modell által szolgáltatott eredményekhez tartozó mérőszámok és mutatók meghatározása, és a tesztelést végrehajtók kijelölése. A modellértékelés validációs, és külön tesztelési fázist is tartalmazhat. Validáció során a modellnek ismételtén a tréning adatbázistól eltérő adatkészleteket adnak bemenetként, melynek során az algoritmus finomul. Ehhez képest a tesztelési fázis egy harmadik, ismét eltérő adatkészletet alkalmaz, melyet még soha nem „látott” az algoritmus. Ezen a harmadik, tesztelési adatkészleten elért teljesítőképességet kell a minősítési jelentésbe bevenni.

Minden felügyelt tanulási technikához ki kell választani azt az alapigazságot (tényleges állapotot leíró jellemzőt), melyre vonatkozó döntések tekintetében az AI modellt értékelni kívánjuk. Ez potenciálisan tartalmaz képességjelzőket és/vagy észlelési eredményeket is, ahogy az a korábbi „Adatok etikai kérdései” részben már tárgyaltuk. Még ha meg is határoztuk az alapigazságokat, etikai nehézségek merülhetnek fel. Például magas szintű

bizonytalansággal járó klinikai helyzetekben az emberi véleményező hajlamos az óvatosság irányában tévedni. Ezt mutatja az a kísérleti tanulmány, melyben nehezen differenciálható jóindulatú és rosszindulatú bőrelváltozások szerepeltek: itt a véleményező orvosok rendszeresen túldiagnosztizálták a rosszindulatúnak vélt eseteket [62].

A tesztelési folyamat során ellenőrizni kell az adatokat: jól illeszkednek-e a majdani alkalmazási környezet adataihoz? Szükséges lehet kiinduló statisztika készítése a tréning és tesztelési adatbázisról, hogy a például megértsük a kérdéses kór előfordulási eloszlását. A döntéshez ún. konfúziós mátrix tartozik:  $TN + TP + FP + FN$  (TN: true negative, valós negatív; TP: true positive, valós pozitív; FP: false positive, téves pozitív; FN: false negative, téves negatív), melyet gyakran alkalmaznak bináris (kétértékű) besorolási problémák esetén (lásd az alábbi 1. ábrát)

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

  

		Prediktív osztály	
		Igen	Nem
Tényleges osztály	Igen	TP	FN
	Nem	FP	TN

1. ábra: A prediktív (véleményezett) és tényleges osztályokat szemléltető konfúziós mátrix  
 A táblázat mutatja, milyen gyakori a helyes és a téves besorolás hozzárendelés

Forrás: Wikipedia (By Oritnk CC BY-SA 3.0, <https://en.wikipedia.org/w/index.php?curid=36792478>)

A modell teljesítőképességének alapos vizsgálatához különféle osztályok/csoportok feldolgozása szükséges. Például meg kell határozni a konfúziós mátrixot az általános, teljes népeiségre, majd külön a női és férfi csoportokra, hogy a nemek okozta esetleg jelenlévő adatorzulás megállapítható legyen. A modell értékelésénél alkalmazott adatbázisnak demográfiai paritást kell mutatnia, melynél minden alanyuk egyforma esélye van a kiválasztásra, továbbá prediktív paritással is rendelkeznie kell, ahol az alanyok predikciójának egyenlő esélyei vannak. Gyakorlatban nehézségekbe ütközhet mind a négy konfúziós mártixelem egyensúlyát elérni a csoportok között. Ezért célszerű lehet csak a mátrix egyes elem párosainak egyensúlyát vizsgálni, például a FP és FN elemekét. A téves pozitív arány egyensúlyának egyenlőnek kell lenni a népeiségcsoportok között, mert ez biztosít az összes résztvevő számára azonos kezelési esélyeket. Új mérőszám vagy mutató, például kiegyenlített esély elérésének kritériuma új szempontú modell-tesztelést tesz lehetővé, hogy az eleget tegyen a téves pozitív és téves negatív arányok kitűzött kritériumának.

A diagnosztikai problémákkal gyakran szembesülő radiológusok tisztában vannak a „túldiagnosztizálás” vagy „aluldiagnosztizálás” veszélyeivel, amiket figyelembe vesznek a döntéshozatalban. Például, ha röntgenfelvétel leletezésekor kétséges észleléseik vannak abdominális szabad légtartalommal kapcsolatban, akkor inkább eltekintenek a specificitás



követelménytől, nehogy a pneumoperitoneum elmulasztott diagnosztizálásának klinikai következményei fenyegezzenek. Mivel a klinikumban fellépő kedvezőtlen vagy ártalmas mellékhatások, egyéb szociális faktorok nehezen felmérhetők és nehezen modellezhetők, az etikus algoritmusok tervezőinek arra kell törekedni, hogy a modell teljesítőképességét - az egyszerű találati pontosságon túl - tényleges alkalmazási körülmények között mérjék és értékeljék. Ez viszont gyakran precízebb statisztikai értékelést igényel, mint a konfúziós mátrix TP, TN, FP és FN elemeiből nyerhető eloszlási görbe alatti terület (AUC: Area Under Curve) egyszerű megállapítása.

## Átláthatóság, értelmezhetőség és indokolhatóság

Az átláthatóság, értelmezhetőség és indokolhatóság vezet a páciens és az egészségügyi szolgáltató közötti bizalom kialakulásához. Ha a radiológus hibázik, akkor szeretnénk tudni, miért következett ez be – részben azért, mert azt is tudni szeretnénk, hogy a hibázás megbocsátható-e vagy sem. Szeretnénk kideríteni, hogy a hibázás mögött figyelmen kívül, netalán rosszindulat áll, esetleg valami egyéb tényező. Hasonlóan, amikor egy AI algoritmus hibázik, vagy döntésével hozzájárul egy téves beavatkozáshoz, kedvezőtlen klinikai eseményhez, a radiológusnak képesnek kell lenni az okok megállapítására.

Némely, a radiológiában elterjedten alkalmazott AI eszköz, például a mesterséges neurális hálózat működésében „fekete doboz” jellegű. Az ilyen eszközökről már kiderült, hogy nehéz megérteni, miért és hogyan hoznak adott döntéseket. A páciensellátásban, ahol a döntéseknek potenciálisan nagy kockázatuk van, a „fekete doboz” megközelítés elfogadhatatlan.

Az értelmezhetőség az AI modell működésének megértére vonatkozik. Az indokolhatóság, vagy megmagyarázhatóság az a tulajdonság, amelynek révén az ember képes felfogni: mi is történik valójában az AI modell döntéshozatala során. Az indokolhatóság azt is felfoghatóvá teszi, hogy az algoritmus miért hozott egy adott döntést, és tartalmazza annak megértését: miért sikeres a modell egyes esetekben, és miért nem működik megfelelően más helyzetekben. Az indokolhatóság magában foglalja az algoritmus struktúra technikai szempontjainak megértését is, és megmagyarázza: miként prezentálja az AI rendszer az eredményeket a páciensnek [63]. Komplex, hálózatba szervezett AI modelleknél a kimeneti eredmények felhasználói lehetnek maguk is AI alrendszerek, a döntéshozatali blokkok kaszkádba szervezett láncolatának további szintjein. Az indokolhatóság, a megmagyarázható AI (XAI: eXplainable Artificial Intelligence) fontos kutatások középponti eleme, melyeket olyan USA szövetségi szintű ügynökségek támogatnak, mint a DARPA (Defense Advanced Research Projects Agency) [64].

Az átláthatósághoz a modelleknek látható szerkezetűnek, megérthetőnek kell lenni a szemlélő számára. Hogy az átláthatóság milyen mértékű kell legyen, az vitatható. A túlzott átláthatóság a rendszert sérülékenyebbé teheti a rosszindulatú támadásokkal szemben, és felfedhet személyes intellektuális tulajdon elemeket.

Az Európai Unió GDPR szabályozásának meghatározása szerint az automatizált döntéshozatali rendszereknek jelentős hatása van a személyre, ezért annak hozzájárulása

nélküli felhasználásuk nem engedélyezett [17, 65]. Meghatározza, hogy ezen túlmenően a személynek joga van magyarázatra: hogyan jött létre az automatizált döntés, és mik annak következményei [66]? Ez úgy értelmezhető, hogy az AI révén létrejött döntések racionálisan, emberi értelem számára felfoghatóan megmagyarázhatók kell legyenek [67].

A radiológus társadalomnak szükséges van irányelvekre, melyek segítségével magyarázhatja, tesztelheti, és egyéb módon elemezheti az AI rendszerek működését. Az irányelvek a klinikai alkalmazások széles körét kell figyelembe veyék. Például, egy MRI szkennerekbe épített, a vizsgálati időt csökkentő AI eljárás különböző hatással lesz a különböző páciensekre, és természetesen mások a technikai buktatói, mint egy képelemző AI rendszernek. Vajon a GDPR értelmében a páciens beleegyezését igénylő, „jelentős hatásnak” számít-e az MRI vizsgálati idő csökkenése? Ugyanígy kérdés: jelentős hatást képvisel minden képelemzésen alapuló AI döntés?

Nem világos, hogy a radiológus milyen mélységig kell egy AI megoldás belső működését felfogja és megértse ahhoz, hogy alkalmazhassa azt a páciensellátásban. Ugyanígy nem tiszta, mennyire transzparensnek kell lenni az AI fejlesztőinek és forgalmazójának termékük belső működése tekintetében. Elegendő, ha az AI termék előállítója átfogó, felszínes magyarázattal szolgál a termékkel kapcsolatban, vagy minden egyes radiológus részletekbe menő tudással kell rendelkezzen a modellről, melyet tudást világosan képes továbbadni a páciensnek? Mi számít adekvát, vagyis elegendően jó magyarázatnak?

## Megismételhetőség

Az AI modellek megismételhető eredményeket kell produkáljanak, azaz a modellnek ismétlés esetén azonos vagy javuló eredményt kell szolgáltatni ugyanazokkal a bemeneti adatokkal. Míg az automatizált rendszerekre ez természetesnek hangzik, ellentétes a humán megfigyelők viselkedésével, akik döntéseikben megfigyelők közötti és megfigyelőn belüli variabilitást is mutatnak. A gépi tanulási (ML) rendszerekkel szembeni minimum követelmény, hogy legalább a humán véleményezővel azonos teljesítményt nyújtsanak. A megismételhetőség függ a problémától is, és a variabilitás mértéke is függ a specifikus feladattól.

## Algoritmus adatorzulás

A számítógéppel támogatott döntéshozás a döntés alapjául szolgáló adatok minőségétől és pontosságától függ. A fentebbi fejezetekben részletesen leírtuk, hogyan hat ki a kiinduló adatkészletek adateltérése, adatorzulása az eredményre. Hasonlóan az emberi döntéshozók is csak ismereteik, korábbi tanulási folyamatuk alapján tudják döntéseiket meghozni.

Az AI orvosi képalkotásban történő alkalmazásával kapcsolatos etika érvényre juttatása a modellek és az adatok folyamatos verifikálásától függ. A már telepített és alkalmazott AI modellek monitorozást igényelnek, majd újrachangolásukra van szükség, ha ellenőrzésük adatorzulást tárt fel, és azonosította annak forrását. Lehetőség van az érintett felek

bevonására, hogy auditálhassák az AI modelleket adattorzulás szempontjából. Javasolt a felhasználók visszajelzéseit és panaszait monitorozó, szabványosított mechanizmus kidolgozása. Gondoskodni kell arról, hogy a modellek üzembe helyezése előtti tréning folyamat adatkészlete illeszkedő legyen a végleges alkalmazás adataihoz. Az AI teljesítőképességét jellemző mérőszámokat és mutatókat alaposan tesztelni kell, azokat a valós alkalmazási körülmények közötti kell a teljesítmény felmérésére kell alkalmazni.

## A gyakorlat etikája

### Számítógép és emberi interakció – a döntési körben maradó humán tényező

Az IEEE (Institute of Electrical, and Electronics Engineers) szervezete legutóbbi publikációiban meghatározta: az autonóm és intelligens rendszerek „mindig is alárendeltek kell legyenek az emberi megítélésnek, felügyeletnek és vezérlésnek” [13], aminek felelőssége esetünkben végső soron a radiológusokra hárul. Ez tekinthető az AI rendszerek egyfajta megközelítésének, ugyanakkor elmulasztja elismerni az autonóm AI eszközök potenciális képességeit és jelentős előnyeit.

Az orvos-páciens reláció a bizalmon alapul. Ahogy nő az orvoslás komplexitása, ez a bizalom az egyéni orvos szolgáltatóról áttevédik nagy egészségügyi intézményekre. Ahogy az intézmények és az önálló orvosok is egyre fokozódó mértékben használnak AI eszközöket, az átláthatóság fenntartása egyre fontosabb lesz a bizalom fenntartása érdekében [7].

Határozottan etikusnak számít, ha transzparenciát biztosítunk a páciensek és minden résztvevő számára - különösen akkor, ha a döntéseket algoritmusok hozzák, vagy legalábbis erős befolyásuk van a döntésekre. Elgondolkodtató helyzeteket eredményezhet például olyankor, amikor a pácienssel megosztott felelősségű megbeszélés másik döntéshozatali szereplője AI eszköz.

Az AI eszközöket alkalmazó radiológusoknak és az intézményeknek átláthatóságot kell biztosítani a páciensek számára. Ennek során a páciens tájékoztatni kell arról: mit történik vele és adataival. A páciensekben tudatosítani kell a következőket:

- Milyen módon felügyeli emberi tényező az AI eszközök döntéseit?
- Miként alkalmazzák az AI eszközöket a diagnózis felállításában, az orvosi javallatok létrehozásában:
- Milyen ellenőrző módszereket használ az intézmény az AI eszközök felmérése, validálása és monitorozása céljából:

Az etikai áttekintés túl kell terjedjen az AI eszközök végfelhasználóin. Előírás, hogy az AI rendszerek fejlesztői, adaptálói és fenntartói mind kövessenek meghatározott etikai elveket [13]. Ugyanígy az AI alkalmazások gazdasági előnyeiben érdekelt szereplőket (például

orvospraxis-menedzserek, egészségügyi finanszírozók) is be kell vonni az etikai megfontolásokba és döntéshozatalba. Az egészségügyi szolgáltatók máris elkezdtek az AI módszerek érzékelhető jótéteményének és előnyeinek reklámozását, hogy ezzel magukhoz vonzzák a pácienseket. Azonban megállapítható: az AI rendszereket nagyon könnyű úgy programozni, hogy a felhasználókat meghatározott minőségi- és teljesítménymutatókat hozó klinikai műveletekre ösztönözzék, vagy a profitot növeljék – anélkül, hogy ez szükségszerűen bármi előnnyel járna a páciens szempontjából. Ahogy a hálózatba kapcsolt rendszerek egyre összetettebbé és dinamikusabbá válnak, egyre nehezebb a felelősség megállapítása egyes AI komponenseknél, a gép-ember közötti felelősség-megosztásról nem is beszélve [68].

Hogyan tartható fenn ilyen körülmények között az átláthatóság? Bizonyára szükség van bizottságokra, felügyeleti szervekre és munkacsoportokra, melyek feladata az AI bevezetésének, klinikai alkalmazásának és hatásainak tüzetes elemzése. Hasznos a folyamatos orvosi-szakmai továbbképzés (CME: Continuous Medical Education), melynek révén az egyes radiológusok hozzájárulhatnak az általuk a napi klinikai rutinban alkalmazott AI eszközök monitorozásához. Az AI modellek korlátjairól, gyengeségeiről és hibáiról gyűjtött, összesített és közzétett információk jól szervezett rendszere biztosíthatja az átláthatóságot, a minőségbiztosítás folyamatosságát és a fejlődést.

Meg kell határozni azokat a részfeladatokat és döntési köröket, melyek nem bízhatók AI modellekre, így biztosítható az emberi felügyelet fennmaradása, és megelőzhetők a pácienseket fenyegető potenciális veszélyek. Hogy a felügyeleti bizottságok és fórumok rendelkezzenek-e formális törvényhozási joggal, azt mindenütt helyi szabályozás határozza meg. Lehet, hogy elegendő, ha ezeknek a testületeknek a jogköre szakmai szervezeteknél, egészségügyi intézményeknél és akadémiai orvostudományi struktúráknál marad (amennyiben ezek az entitások bírják munkatársaik, a páciensek és a közvélemény bizalmát). Az a jogi kérdés, hogy az autonóm AI eszközök kezelése eltérjen-e a közvetlen emberi felügyelet alatt álló eszközökétől, még megfontolás és döntés tárgya [69].

## Oktatás

A mesterséges intelligencia eszközök a radiológusok és egyéb radiológiai szerepkörök kiváltása helyett inkább új, az eddigiektől eltérő jártasságot és szakértelmet igényelnek majd. Ez egyedülálló lehetőséget ad a radiológiai munkafolyamat alapvető elemeinek átértékelésére, melynek révén meghatározhat az emberi és AI eszközök munkavégzéshez szükséges, optimális kombinációja. A radiológiának még kutatásra és specifikus képzési támogatásra van szüksége; a radiológusok és páciensek protokollokat igényelnek az új, megosztott döntéshozás paradigmáinak bevezetéséhez. A képzés egy részének a majd rendelkezésre álló AI eszközök alkalmazásának gyakorlati kérdéseire kell összpontosítani. Másrészt figyelni kell a képzés során az új AI módszerek bevezetésével felmerülő etikai problémákra is. Az etikus adatgyűjtés és adatfelhasználás, és az etikus algoritmus-fejlesztés biztosításán túlmenően (mindkét folyamatot aránylag kisszámú szakember irányítja és ellenőrzi majd) gondolni kell az AI algoritmusok mindennapos, etikus használatával kapcsolatos felelősségre is. Ebben minden orvos és radiológus érintett, akinek a munkája AI elemeket tartalmaz. Az AI eljárások etikus használatát legjobban úgy lehet biztosítani, hogy a nap mint nap ilyen eszközöket alkalmazó orvosok közegeben tudatosítjuk az AI eszközök

használatának morális kockázatát és felelősségét. Minél képzetesebbek a radiológusok, annál kevesebb lesz az akaratlan vagy egyéb tévedés és hibázás ezen a területen.

## Automatizálási elfogultság

Automatizálási elfogultság az emberi tényező hajlama a gépi úton létrejött döntések előnyben részesítésére, melynél a gépi döntésnek ellentmondó adatok, vagy ellentétes humán döntések figyelmen kívül maradnak. A szakirodalom számos példát tartalmaz olyan „automatizálási elfogultság” esetekre, melyek akkor lépnek fel, amikor emberek felügyelnek gépi döntési rendszereket, főként nagy bonyolultságú helyzetekben [70]. Ez az elfogultság a gépi döntéshozatallal való visszaélésekhez vezet, melynek következményei a monitorozás és felügyelet hiányossága, felesleges gépi döntéstől való függés és indokolatlan vak bizalom [71]. A klinikai támogató rendszerek körében is tapasztalható automatizálási elfogultságról is létezik alapos elemzés [72].

Az automatizálási elfogultság figyelmen kívül hagyás, és túlzott alkalmazás típusú hibához egyaránt vezethet. Figyelmen kívül hagyásos hiba az, amikor a rendszer emberi szereplője elmulasztja az AI rendszer hibáinak észlelését, vagy nem törődik az ilyen hibákra utaló jelekkel. Ilyen hibákra hajlamosít az a környezet, ahol a radiológiai vizsgálat és leletez nagy sebességgel történik, valamint gyors a kapcsolódó döntéshozatal is. Ezt a jelenséget bonyolítja az olyan AI döntéshozatal, amely az emberi értékelő számára nem észlelhető, rejtett tényezőknél és tulajdonságokon alapul. A másik típusú hibára, a túlzott alkalmazásra példa az, amikor a radiológus tévesen elfogadja, vagy érvényre juttatja a gépi döntést, esetleg annak ellenére, hogy ellentétes bizonyítékok állnak rendelkezésre.

A radiológia már szembesült az automatizálási elfogultság jelenségével az ún. számítógéppel támogatott észlelés (CAD: Computer Aided Detection) algoritmusok használata során, például ahol mammográfiai szűrések képeit értelmezik és leletezik segítségükkel. Az ilyen CAD módszer az Amerikai Egyesült Államokban az FDA által jóváhagyott eljárás, és finanszírozza a MediCare egészségbiztosító. Tanulmányok bizonyították, hogy a CAD használata összefügg a mammográfiai szűrések csökkent pontosságú értékelésével, melyeknél megnő a visszahívás és a biopszia részaránya, egyes radiológus csoportoknál még a vizsgálatok szenzitivitása is csökkenést mutatott [74]. Feltételezések szerint a szenzitivitás csökkenés összefüggésben lehet a CAD eredményektől való túlzott függéssel, az ilyen döntésekbe vetett túlzott bizalommal. Míg az AI alapú CAD algoritmusok kísérleti körülmények között sokkal ígéretesebbek, mint a hagyományos CAD észlelési eljárások, az még nem világos, hogyan befolyásolják majd a humán-gépi AI interakciók a vizsgálatok pontosságát vagy hatékonyságát a tényleges klinikai gyakorlatban. Erre vonatkozó célzott vizsgálatot jelent például az FDA engedélyezési eljárás, mely a biztonságot és a hatékonyságot is elemzi. Egyfajta „nem szigorú szabályozás” (soft governance) is hasznos lehet: valószínűleg az AI (és egyéb) termékek nem kelnének el a piacon széles körben akkor, ha nem bizonyított, hogy eleget tesznek az elfogadott standardoknak - akár megköveteli ezt a helyi törvényi szabályozás, akár nem [75].

## Páciens preferenciák

Egy 2017-ban az Amerikai Egyesült Államokban végzett felmérés szerint az amerikai felnőtt lakosság 65%-a kényelmetlennek érezné, ha az orvosi diagnózis létrehozásának feladatát AI eszközökkel rendelkező számítógépre bíznák [76]. További kutatás szükséges ahhoz, hogy megértsük: mikor és hogyan lesznek hajlandók a páciensek elfogadni a gép által hozott döntéseket, mikor bíznak majd az ilyen döntésekben, és mikor válik szükségessé mindez?

A radiológiai szakmának figyelembe kell venni a páciensek kollektív kívánságait az AI eszközök alkalmazásával kapcsolatban. Ezek a kívánságok nem biztos, hogy egybevágnak az AI modelleket létrehozó logikával. Például az autonóm gépjárművek (AV: autonomous vehicle) döntéshozatali rendszereivel kapcsolatos tanulmányok azt jelzik, hogy az emberek helyeselnének egy olyan haszonelvű AV járművet, mely a benne ülőket feláldozná a nagyobb előny érdekében, ha dönteni kellene egy gyalogos elgázolása, és a bent ülők épségben maradása között – de inkább azt szeretnék, ha mások vennének ilyen kocsikat. Másrészt, az emberek szívesebben utaznának olyan AV járműben, amelyik a benne ülőket minden áron védelmezi [77]. Bonyolítja a helyzetet, hogy az AI-vel kapcsolatos normák kulturálisan eltérők világszerte [78], ami arra utal, hogy egységes szabályozás gyakorlatilag lehetetlen.

A nyilvánosság hasonló ambivalenciája valószínűsíthető a mesterséges intelligencia radiológiában történő alkalmazásának kérdésében is. Vajon elfogadja-e a társadalom az AI részvételével működő radiológia tökéletlenségeit, a nagyobb előny lehetőségének érdekében? Vagy az egyén saját magára és szerettei vonatkozó döntéseiben sokkal kisebb toleranciát mutat majd az ilyen tökéletlenségekkel szemben? Ha például az orvosi képzést tisztán protokoll alapú, és algoritmusok által értékelt lenne, vajon maradna-e hely a józan ész alkalmazásának? Marad-e hely az egyéni és a lakosság sugárterhelés-kockázat, és azzal szemben az individuális páciens vizsgálati elvárások mérlegelésének? Ha elismerjük az AI által működtetett radiológia tökéletlenségét és gyors fejlődését, vajon a társadalom is elfogadja-e majd, mert esetleg kevésbé költséges, vagy kevésbé munkaigényes, mint a hagyományos, hús-vér orvosok által működtetett radiológia?

## Nyomonkövethetőség

A nyomon követhetőség a dolgok összekapcsolásának, és a kapcsolódás követésének képessége. A követhetőség döntő fontosságú a páciensek és az egészségügyi szolgáltatók ilyen rendszerekbe vetett bizalma miatt. A nyomon követhetőség lehetőséget ad arra, hogy a nem elvárt módon működő rendszerkomponenseket azonosíthassuk, működtethető legyen a minőségbiztosítás, és alkalmazhatók legyenek a szükséges korrekciós lépések.

A követhetőség koncepciója a teljes szoftver-létrehozási folyamatra vonatkozik. Egy radiológiai AI esetben például a leletsablon előírt diagnózis mezője (melynek tartalma lehet X betegség megléte vagy hiánya) hozzáköthető egy AI modellhez, amelyik ezt a besorolást generálja. Ha ez a kapcsolódás megvan, nyomon követhető a reláció, mellyel a besorolás létrejötté ellenőrizhető. Ugyanígy, a besorolás visszavezethető az AI modellre, amely a döntést létrehozta. A nyomon követhetőség a szoftver-tesztelésben a lefutás előre és hátra

irányú, lépésenkénti elemezhetőségét jelenti, melyben általában kontrollált teszteseteket alkalmaznak. Tesztelhető az AI modell ellenőrzött környezetben is, hogy megállapítható legyen: eleget tesz-e a specifikációknak? A teszt körülmények közötti nyomon követés eredményeit mátrixba szerezve megállapíthatók a követelmények teljesüléseinek összefüggései.

## Mesterséges intelligencia és munkaerő-erózió

Az AI-vel kapcsolatos legelterjedtebb előítéletek egyike az, hogy bevezetésével emberek vesznek el állásukat [75]. A radiológusokat is érintheti ez a lehetőség, és a vele kapcsolatos félelem is. Ez olyan viselkedéshez és gyakorlathoz vezethet, ami az emberi orvoslás szerepköreit és relevanciáját igyekszik fenntartani a jövőben, függetlenül attól, hogy az ilyen közvetlen emberi közreműködés fenntartása a közjó szempontjából végül is hasznos-e, vagy sem.

Az AI alkalmazásának etikai kérdéseit környező jelenlegi viták középpontjában a következő vélekedés áll: az emberi tényező kulcsszerepe az AI alkalmazásánál az, hogy megakadályozza az új technológia bevezetésének esetleges negatív következményeit. Naivitás lenne figyelmen kívül hagyni azt a lehetőséget, hogy ennek során az ember saját érdekeit igenis szem előtt tartva cselekszik. Tudomásul kell venni, hogy az ember mélyen gyökerező érdeke, hogy ne váljon teljesen feleslegessé egy kialakulóban lévő technológia következtében. Továbbá természetes, hogy radiológusok készletet éreznek majd relevanciájuk és érdekeik megvédésére, ha egy lehetséges jövőben a hierarchiában elfoglalt helyüket veszélyeztetik vagy akár megszüntetni fenyegetik holmi informatikai szakemberek, vagy egyéb, nem hagyományos egészségügyi szereplők. Nemcsak a társadalmat és a pácienseket szükséges megvédeni a „robotvezérelt radiológia” veszélyeitől, de ezzel szemben arra is szükség van, hogy a radiológusok és egyéb orvosok önérdeke ellen hassunk, ha ez az önérdek konfliktusba kerülne a közjó követelményeivel.

Egyelőre egyszerűen nem tudhatjuk, milyen lesz a páciensek és az erőteljesen AI-t alkalmazó radiológia kölcsönhatása. Lehet, hogy az új technológia egyes elemeit üdvözlés fogadja majd, más vonatkozásai pedig félelmet váltanak ki, és jelentős ellenállásba ütköznek. A jövőre vonatkozó egyik lehetséges jellemzés a „liberal eugenics”, amit talán a kiváltságos réteg létrejöttéként értelmezhetnénk. Egy ilyen elképzelt jövőben a lakosság kis, speciális tudással és adathozzáféréssel rendelkező része ezeket az előnyöket öncélúan használja ki. Például előnyt élvezhetnek egy költséges radiológiai szűrést végző AI eszköz alkalmazásából [79].

## Erőforrás egyenlőtlenség

Az AI alkalmazása nagymennyiségű adathoz történő hozzáférést, adatkezelési képességeket és gyakorlatot igényel. Hatalmas számítástechnikai apparátusra van szükség a komplex öntanuló rendszerek trenírozásához és menedzseléséhez. A kisebb vagy erőforrásokkal gyengébben ellátott kórházak és kutatóintézetek esetleg nem rendelkeznek ezekkel a lehetőségekkel és eszközökkel. Majdnem biztos, hogy az AI alkalmazások valamely része a

fejlesztő saját tulajdonú terméke lesz, melyet nagy akadémiai vagy privát egészségügyi szolgáltatók, biztosító társaságok és professzionális nagy szoftverházak fejlesztenek és birtokolnak. Előfordulhat, hogy ezek nem mindegyike rendelkezik túl nagy hagyománnyal és tudással a radiológiában. Mindezek a veszélyek súlyosbítják a kutatókapacitás és a szolgáltatások egyensúlyának problémáit.

Az AI modellfejlesztők számára kellő pénzügyi forrás és ösztönzés kell rendelkezésre álljon a folyamatos fejlődés támogatásához, de gondolni kell az erőforrásokkal szerényen ellátott közösségekre is. Megfizethető áron, állami és privát alapokból elérhetővé kell tenni a lakossági egészségügyi szolgáltatásokban bizonyítottan jól teljesítő modelleket és AI megoldásokat. Ugyanakkor tudatosítani kell a radiológusokban és az AI megoldások egyéb felhasználóiban azt, hogy az erőforrásokkal és lehetőségekkel kevésbé rendelkező közösségek vonatkozásában egyes részpopulációk jelentős alulreprezentáltsága fordulhat elő az algoritmusok tréningjénél és tesztelésénél. Az ilyen adatorzulások tudatosítása segítheti a modellek bevezetésénél jelentkező problémák felismerését és orvoslását. Ennek érdekében az AI modellek teljesítőképességét monitorozásával megbízott tanácsadó csoportok és intézményi szakbizottságok összeállításánál gondolni kell arra, hogy mindenféle háttérű és szakértelmű szakember szerepeljen bennük a kellő képviselési egyensúly megvalósításához.

## Felelősség

A felelősség problematikájának egy érdekes felvetése, hogy maga a mesterséges intelligencia eszköz egyáltalán felelőssé tehető-e működése és döntései ügyében, és ha igen, milyen módon? Ez elsősorban jogi kérdés, bár etikai és erkölcsi szempontok is befolyásolják a kérdésre adandó választ. A dolgok jelen állása szerint egyelőre a végső felelősség és felelősségre vonhatóság az emberre vonatkozik [68].

Az orvosi gyakorlatban történő AI modell alkalmazás etikáját vizsgálva meg kell válaszolni a kérdést: hol a felelősségre vonhatóság akkor, ha az alkalmazásban gyengék vagy kedvezőtlenek a páciens eredmények? Jelenleg az orvosok, köztük a radiológusok tarthatók felelősnek, ha az egészségügyi standardok egy szolgáltatásnál nem teljesülnek. Ahol az emberi döntések támogatására AI eszközt használnak, ott valószínűleg továbbra is a radiológus marad a felelős. Mivel az AI modellek jelentős mennyiségű adatot, köztük az ember számára nem észlelhető és így fel sem fogható információt tartalmaznak, felmerülhet a kérdés: a gyenge vagy rossz eredményekért tényleg teljes mértékben az orvost kell-e felelősségre vonni, vagy a felelősséget részben, vagy teljesen át lehet és kell hárítani azokra, akik az AI terméket előállítják, piacra viszik és eladják?

Átláthatóság szükséges a radiológiai AI alkalmazásoknál ahhoz, hogy megállapíthassuk: a modell valamely vétkes hibája hozzájárult-e a gyenge vagy rossz páciens eredményhez? Vajon a kórház, vagy a modellt használatba vevő egészségügyi rendszer-e a felelős ez esetben? Az is kérdés, mi van akkor, ha a gyenge páciens eredmény okozója a radiológus, aki az AI modell kimeneti eredményének ellenében, de a legjobb tudása szerinti döntést hozta? Manapság a radiológus felelősségének kérdése főként a lehetséges hanyagsághoz kapcsolódik: elvárható ésszerűséggel és alaposággal járt-e el az orvos? Az autonóm gépi működések korszakában, amikor egyáltalán nincs ember a döntési láncban, esetleg az kerül



majd a kérdés középpontjába, hogy a számítógép olyan jól teljesített-e, ahogy kellett volna neki [18, 69]?

## Érdekkonfliktusok

Az érdekkonfliktus meghatározása: „bizonyos körülmények esetén fellépő kockázat, ahol egy elsőrendű érdeket figyelembe vevő szakmai döntést vagy cselekvést helytelenül befolyásol egy másodlagos érdek” (COI: Conflict Of Interest) [80, 81]. Az olyan, éppen most születő és fejlődő piacokon, mint a radiológiai AI alkalmazások területe, várható és teljesen természetes, hogy a páciensellátásban is szereplő radiológusok és egyéb szakemberek esetenként pozíciót vállalnak AI startup cégekben, vagy később a megerősödő, az egészségügyi szolgáltatások piacán nyomuló vállalatokban. Hasonlóan ahhoz, amikor egy klinikai tanulmánynál a gyógyszer hatását értékelő személy anyagilag érdekelt az adott gyógyszer sikerében, a radiológusoknál is előfordulhat AI eszközökkel kapcsolatos érdekellentét. Ezekre a problémákra megoldásként a transzparencia, nyilvánosságra hozatal, intézményi felügyelet, tulajdonrésztől való megfosztás és hasonló ellenintézkedések jöhetnek számításba.

Egyes esetekben az orvos vagy egyéb egészségügyi adminisztrátor személyes érdekellentéte (rangja és beosztása révén) jelentős mértékben intézményi érdekkonfliktusnak is számít. Az Amerikai Egyesült Államokban ezt a problémát az „American Association of Medical Colleges” szakmai szervezet az alábbiak szerint határozza meg: „... egyes személyek hivatali pozíciója olyan közvetlen, erős hatalommal, tekintéllyel és felelősséggel járhat a vezetői feladatokkal és a kutatási programokkal kapcsolatban, hogy az illető személy pénzügyi érdekeltisége, és az intézet kutatási alanyainak érdekei közötti ütközés már intézményi érdekkonfliktusnak tekinthető [82].” Intézményi érdekkonfliktus esetén az intézménynek kreatívnak kell lenni, hogy például független külső felügyeleti auditálási lépésekkel előzze meg a társadalom bizalmának elvesztését.

Az egészségügyben fellépő érdekkonfliktus esetén akár az egyénnek, akár intézménynek ügyelnie kell arra, hogy az átláthatóság biztosításával, publicitással megfelelően kezelje a kérdést [83, 84]. Az egészségügyi AI alkalmazásokkal összefüggő pozíciókban lévő személyek (páciensek adatait harmadik, nem egészségügyi félnek kiszolgáltató felelősök, AI beszerzési ügynökök, AI modellek klinikai implementációját végzők) különös gonddal kell kezeljék esetleges érdekütközéseiket. Egyes esetekben az ilyen konfliktusok az ilyen személyeknél akár tevékenységről történő lemondást, vagy pozícióból való visszahívást is szükségessé tehetnek.

## Következtetések

A mesterséges intelligencia (AI) határozott potenciált jelent a radiológia fejlődése, a páciensek ellátásának javítása területén, ugyanakkor költséghatékonyabbá teheti az orvosi képző eljárásokat. Bevezetése azonban komplex etikai és szociális kérdéseket vet fel. Jelen közlemény szándéka és célja a kollektív, a mesterséges intelligencia radiológiában történő alkalmazásával kapcsolatos szakmai egyeztetés.

Minden, az AI alkalmazásában érintett résztvevő alaposan meg kell értse az AI lényegét, meg kell tudja becsülni: milyen kapcsolódó veszélyek jelentkezhetnek. Ezeket a veszélyeket átláthatóvá kell tenni, és minden el kell követni a kockázatok csökkentése, a lehetséges károk megelőzése érdekében. Különösen a radiológusokra hárul nagy felelősség, hogy helyesen felmérjék az AI alkalmazások előnyeit és hátrányait, felhívják a páciensek és egészségügyi beruházók figyelmét a kockázatokra, valamint folyamatosan figyeljék az AI eszközök teljesítőképességét, viselkedését és hatásait a veszélyek megelőzése érdekében. Még ha megfelelően etikus is bevezetése, az AI óhatatlanul szociális és gazdasági kihatású változásokat eredményez majd. A változások legtöbbször pozitív lesz, de némely bizonyulhat kedvezőtlennek is.

Az AI alapvetően megváltoztatta felfogásunkat a radiológiai vizsgálatokról és az ahhoz kapcsolódó adatokról, az adatok alkalmazásáról, de az adatokkal kapcsolatos esetleges visszaélésekről is. A radiológus kötelessége felfogni az AI alkalmazás problematikáját, de ugyanígy kötelezett megérteni az adatkezeléssel kapcsolatos kérdéseket is. A szélesebb értelemben vett radiológus társadalom erkölcsi kötelessége az összegyűjtött adatok közjó érdekében történő felhasználása, minél több információ kinyerése a páciensről és betegségéről, összességében a radiológia gyakorlatának jobbítása. Etikai kötelességük a páciensektől származó adatok minél hasznosabb feldolgozása éppen azok érdekében, akiktől és akikről az adatokat nyerték.

A radiológusok számára az AI alkalmazások és adatok inkább helyzeti, eseti és kevésbé abszolút értéket jelentenek majd. A szakmai közösség feladata támogatni az AI etikus bevezetését, megfelelő kezelését erősítő hasznos rendszereket és intézményeket, hogy ezzel a mesterséges intelligencia eszközök a klinikai ellátásban, a kutatásban és az üzleti életben egyaránt elterjedhessenek és fejlődhessenek.

A radiológiának már ma el kell kezdeni az AI számára szóló etikai szabályok és eljárásrend kidolgozását. Ezeknek a szabványoknak és szabályozásnak, az etikus AI alkalmazásokra vonatkozó eljárásrendnek a műszaki, klinikai és gazdasági érdekek bonyolult szövevényében kell eligazítással szolgálni, mégpedig kellő morális kritikával. Az etikus AI létrehozása és biztosítása minden résztvevő részéről maximális bizalomra törekvést követel meg. Radiológia-centrikus AI szaktudás és technológiai ismeret egyaránt szükséges az AI produktumok ellenőrzéséhez és engedélyezéséhez. Paradox módon, az imént említettek megvalósításában várhatóan AI eszközök is szerepelnek majd. A verifikáláshoz és validáláshoz szükséges eljárásrend kulcsfontosságú elemei a folyamatos és hangsúlyos transzparencia, a páciensek védelme, az adatkészlet változatok gondos kézbentartása. Fontos a bevezetést követő folyamatos monitorozás, hogy megelőzhetőek legyenek a nem kívánt következmények és minőségi eltérések. Ha ilyenek fellépnek, alaposan ki kell vizsgálni az okokat, és korrekciós intézkedések szükségesek a hibák kijavítására.

A radiológusok egyszerre fejlesztői és alkalmazói a mesterséges intelligenciának, ezzel párhuzamosan ismerkednek az AI etikai vonatkozásaival. Az AI technológia fejlődése és a társadalom erre adott reakciója alig felfogható ütemű és kihatású, még nehezebb e fejlődés kézbentartása. Az etikai aggályok megértése, a kapcsolatos kérdések és megfelelő reagálás egyaránt folyamatosan változnak. A mesterséges intelligencia a radiológia minden

szerepkörére hatással lesz bizonyos mértékben. A páciensek és a közjó érdekében morális kötelezettségünk az etikai kérdések céltudatos és tervszerű elemzése, az adatok etikus felhasználása, a döntéshozó gépezetek létrehozásánál és működtetésének alapos megfontolása, és a kapcsolódó üzleti tevékenység morálisan vállalható alakítása.

## Meghatározások

- AI Artificial Intelligence – mesterséges intelligencia. Számítógépes alkalmazásokra vonatkozó fogalom, melyek olyan módon viselkednek, amiről a legutóbbi időig úgy gondoltuk: csak emberi elme képes ilyen működésre
- ML Machine Learning – gépi tanulás. Egyre több adat feldolgozása során viselkedését folyamatosan változtató, ideális esetben javuló teljesítőképességű algoritmus
- Supervised ML – felügyelt MT, felügyelt gépi tanulás. Adatok és hozzá rendelt jelölők alapján tanuló, változó gépi algoritmus. A klinikai radiológiában a felügyelt ML radiológiai képek és előre meghatározott jelölők, tulajdonságok egymáshoz rendelését megvalósító, repetitív folyamat
- Non-supervised ML – nem felügyelt ML. Ennek során jelölők (azaz ismert válaszok) nélküli adatkészletet elemez a gépi algoritmus, és önállóan csoportosítja a képekben talált tipikus észleléseket azok tulajdonságai alapján. Deep Learning – „mélytanulás” bemenő és kimenő adatok többretegű feldolgozását alkalmazó, nem felügyelt gépi tanulási folyamat
- Neural Networks – neurális hálózat. A „mélytanulás” adaptív, akár struktúráját is változtató algoritmosos változata, mely jól teljesít radiológiai képekkel kapcsolatos döntéshozatalban
- Algoritmus – bemenő adatokat értékeléssel, szelekcióval és döntéshozattal feldolgozó számítógépes eljárás (programkód)
- Modell – AI kontextusban: adatkészleten „trenírozott” algoritmus működésének eredményeként létrejövő modell. Ha új adatkészletekkel trenírozzák ugyanazt az algoritmust, vagy egy adott adatkészlettel más és más algoritmusokat edzenek, újabb és újabb modellek jönnek létre. Ahogy a modell edződik, működése egyre gyorsabb lesz, és csökkenő számítástechnikai ráfordítás szükséges működéséhez, amennyiben a trenírozáshoz használt adatokhoz hasonló adatkészleteket kell feldolgoznia
- Adateltérés, adattorzulás (bias) – az igazságtól való szisztematikus eltérés (statisztikai értelemben vett adat-mintavételi hiba, amikor az adatok nem megfelelően egyenletesen, nem részrehajlás nélkül reprezentálják a valóságot, vagy a jellemezni kívánt halmazt)
- Variancia – az igazságtól való véletlenszerű eltérés (statisztikai értelemben)

megfelelően reprezentált halmazban az elemeknek egy adott tulajdonságtól való random, sztochasztikus eltérése)

## Irodalmi referenciák

1. High P Carnegie Mellon Dean Of Computer Science On The Future Of AI. In: Forbes.  
<https://www.forbes.com/sites/peterhigh/2017/10/30/carnegie-mellon-dean-of-computer-science-on-the-future-of-ai/>. Accessed 19 Feb 2019
2. Kohli M, Prevedello LM, Filice RW, Geis JR (2017) Implementing Machine Learning in Radiology Practice and Research. *Am J Roentgenol* 1–7. <https://doi.org/10.2214/AJR.16.17224>
3. Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine Learning for Medical Imaging. *RadioGraphics* 37:505–515. <https://doi.org/10.1148/rg.2017160130>
4. García-Pedrajas N, Ortiz-Boyer D, del Castillo Gomariz R, Martínez C (2005) Cascade Ensembles. pp 97–115
5. Floridi L, Taddeo M (2016) What is data ethics? *Philos Trans R Soc Math Phys Eng Sci* 374:20160360. <https://doi.org/10.1098/rsta.2016.0360>
6. Mittelstadt BD, Floridi L (2016) The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Sci Eng Ethics* 22:303–341. <https://doi.org/10.1007/s11948-015-1084-9>
7. Li Y (Boni), James L, McKibben J (2016) Trust between physicians and patients in the e- health era. *Technol Soc* 46:28–34. <https://doi.org/10.1016/j.techsoc.2016.02.004>
8. Obermeyer Z, Emanuel EJ (2016) Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 375:1216–1219. <https://doi.org/10.1056/NEJMp1606181>
9. Char DS, Shah NH, Magnus D (2018) Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *N Engl J Med* 378:981–983. <https://doi.org/10.1056/NEJMp1714229>
10. Cabitza F, Rasoini R, Gensini GF (2017) Unintended Consequences of Machine Learning in Medicine. *JAMA* 318:517–518. <https://doi.org/10.1001/jama.2017.7797>
11. European Group on Ethics in Science and New Technologies Statement on Artificial 1095 Intelligence, Robotics and “Autonomous Systems.” European Commission
12. (2017) Association for Computing Machinery 2018 Code of Ethics and Professional Conduct, Draft 3. In: ACM Ethics. <https://ethics.acm.org/2018-code-draft-3/>. Accessed 21 Jan 2019
13. IEEE Global Initiative Ethically Aligned Design, Version 2 (EADv2) | IEEE Standards Association. Institute of Electrical and Electronics Engineers
14. The Montreal Declaration for a Responsible Development of Artificial Intelligence: a participatory process. Montreal Declaration for Responsible AI
15. Bill Text - AB-375 Privacy: personal information: businesses.  
[https://leginfo.ca.gov/faces/billTextClient.xhtml?bill\\_id=201720180AB375](https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375). Accessed 3 Jan 2019
16. Ghosh D (2018) What You Need to Know About California’s New Data Privacy Law. *Harv. Bus. Rev.*
17. General Data Protection Regulation (GDPR) – Final text neatly arranged. In: Gen. Data Prot. Regul. GDPR. <https://gdpr-info.eu/>. Accessed 3 Jan 2019
18. European Parliament (2017) Civil Law Rules on Robotics
19. Protection of Human Subjects
20. Kesner A, Laforest R, Otazo R, et al (2018) Medical imaging data in the digital innovation age. *Med Phys* 45:e40–e52. <https://doi.org/10.1002/mp.12794>
21. Rubin DL, Kahn CE (2016) Common Data Elements in Radiology. *Radiology* 161553. <https://doi.org/10.1148/radiol.2016161553>

22. Shie C-K, Chuang C-H, Chou C-N, et al (2015) Transfer representation learning for medical image analysis. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Milan, pp 711–714
23. Ravishankar H, Sudhakar P, Venkataramani R, et al (2017) Understanding the Mechanisms of Deep Transfer Learning for Medical Images. ArXiv170406040 Cs
24. Balloch JC, Agrawal V, Essa I, Chernova S (2018) Unbiasing Semantic Segmentation For Robot Perception using Synthetic Data Feature Transfer. ArXiv180903676 Cs
25. Goodfellow I, Pouget-Abadie J, Mirza M, et al (2014) Generative Adversarial Nets. In: Ghahramani Z, Welling M, Cortes C, et al (eds) Advances in Neural Information Processing Systems 27. Curran Associates, Inc., pp 2672–2680
26. Perez L, Wang J (2017) The Effectiveness of Data Augmentation in Image Classification using Deep Learning. ArXiv171204621 Cs
27. Torralba A, Efros AA (2011) Unbiased look at dataset bias. In: CVPR 2011. pp 1521–1528
28. Wendler D (2006) One-time general consent for research on biological samples. *BMJ* 332:544–547
29. Smith J (2018) European Data Protection Board - Members. In: Eur. Data Prot. Board - Eur. Comm. [https://edpb.europa.eu/about-edpb/board/members\\_en](https://edpb.europa.eu/about-edpb/board/members_en). Accessed 21 Jan 2019
30. Mission – Medical Data Donors. <http://www.medicaldatadonors.org/index.php/mission/>. Accessed 22 Jan 2019
31. Canada. Supreme Court (1992) *McInerney v. MacDonald*. *Dom Law Rep* 93:415–431
32. (2007) Information Governance of the Interoperable EHR | Canada Health Infoway. Canada Health Infoway Inc.
33. (2014) Public Hospitals Act 34. (2010) Tri-council policy statement. Ethical Conduct for Research Involving Humans. Government of Canada
35. Balthazar P, Harri P, Prater A, Safdar NM (2018) Protecting Your Patients’ Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics. *J Am Coll Radiol* 15:580–586. <https://doi.org/10.1016/j.jacr.2017.11.035>
36. (2018) Algorithms in decision-making. House of Commons, United Kingdom Parliament
37. Mikk KA, Sleeper HA, Topol EJ (2017) The Pathway to Patient Data Ownership and Better Health. *JAMA* 318:1433–1434. <https://doi.org/10.1001/jama.2017.12145>
38. Budin-Ljøsne I, Teare HJA, Kaye J, et al (2017) Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC Med Ethics* 18: <https://doi.org/10.1186/s12910-016-0162-9>
39. Ballantyne A, Schaefer GO (2018) Consent and the ethical duty to participate in health data research. *J Med Ethics* 44:392–396. <https://doi.org/10.1136/medethics-2017-104550>
40. Dubovitskaya A, Xu Z, Ryu S, et al (2018) Secure and Trustable Electronic Medical Records Sharing using Blockchain. *AMIA Annu Symp Proc* 2017:650–659 41. Azaria A, Ekblaw A, Vieira T, Lippman A (2016) MedRec: Using Blockchain for Medical Data Access and Permission Management. In: 2016 2nd International Conference on Open and Big Data (OBD). IEEE, Vienna, Austria, pp 25–30
42. Haug CJ (2017) Whose Data Are They Anyway? Can a Patient Perspective Advance the Data-Sharing Debate? *N Engl J Med* 376:2203–2205. <https://doi.org/10.1056/NEJMp1704485>
43. Mello MM, Lieou V, Goodman SN (2018) Clinical Trial Participants’ Views of the Risks and Benefits of Data Sharing. *N Engl J Med* 378:2202–2211. <https://doi.org/10.1056/NEJMs1713258>
44. Grapevine World Token. <https://grapevineworldtoken.io/>. Accessed 22 Jan 2019
45. All-of-Us Program Overview. In: Us. <https://www.joinallofus.org/en/program-overview>. Accessed 19 Feb 2019
46. Warren SD, Brandeis LD (1890) The Right to Privacy. *Harv Law Rev* 4:193–220.

- <https://doi.org/10.2307/1321160>
47. (2019) Protection of Human Subjects
  48. Mazura JC, Juluru K, Chen JJ, et al (2012) Facial Recognition Software Success Rates for the Identification of 3D Surface Reconstructed Facial Images: Implications for Patient Privacy and Security. *J Digit Imaging* 25:347–351. <https://doi.org/10.1007/s10278-011-9429-3>
  49. Demner-Fushman D, Kohli MD, Rosenman MB, et al (2015) Preparing a collection of radiology examinations for distribution and retrieval. *J Am Med Inform Assoc* ocv080. <https://doi.org/10.1093/jamia/ocv080>
  50. Na L, Yang C, Lo C-C, et al (2018) Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. *JAMA Netw Open* 1:e186040–e186040. <https://doi.org/10.1001/jamanetworkopen.2018.6040>
  51. Carlini N, Liu C, Kos J, et al (2018) The Secret Sharer: Measuring Unintended Neural Network Memorization & Extracting Secrets. *ArXiv180208232 Cs*. <https://doi.org/arXiv:1802.08232v1>
  52. Song C, Ristenpart T, Shmatikov V (2017) Machine Learning Models that Remember Too Much. *ArXiv170907886 Cs*. <https://doi.org/arXiv:1709.07886>
  53. Fairness | Machine Learning Crash Course. In: Google Dev. <https://developers.google.com/machine-learning/crash-course/fairness/video-lecture>. Accessed 19 Feb 2019
  54. (2018) Artificial intelligence and medical imaging 2018: French Radiology Community white paper. *Diagn Interv Imaging* 99:727–742. <https://doi.org/10.1016/j.diii.2018.10.003>
  55. Khullar D (2019) Opinion | A.I. Could Worsen Health Disparities. *N. Y. Times*
  56. Jordan MI, PhD TGD, Storkey A, et al (2008) *Dataset Shift in Machine Learning*, First Edition edition. The MIT Press, Cambridge, Mass
  57. Calvert JS, Price DA, Chettipally UK, et al (2016) A computational approach to early sepsis detection. *Comput Biol Med* 74:69–73. 1194 <https://doi.org/10.1016/j.compbiomed.2016.05.003>
  58. Mao Q, Jay M, Hoffman JL, et al (2018) Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open* 8:e017833. <https://doi.org/10.1136/bmjopen-2017-017833>
  59. Geijer H, Geijer M (2018) Added value of double reading in diagnostic radiology, a systematic review. *Insights Imaging* 9:287–301. <https://doi.org/10.1007/s13244-018-0599-0> 1201
  60. Boers M, Kirwan JR, Wells G, et al (2014) Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol* 67:745–753. <https://doi.org/10.1016/j.jclinepi.2013.11.013>
  61. O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1 edition. Crown, New York
  62. Esteva A, Kuprel B, Novoa RA, et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118. <https://doi.org/10.1038/nature21056>
  63. Gilpin LH, Bau D, Yuan BZ, et al (2018) Explaining Explanations: An Overview of Interpretability of Machine Learning. *ArXiv180600069 Cs Stat*
  64. Explainable Artificial Intelligence. <https://www.darpa.mil/program/explainable-artificial-intelligence>. Accessed 17 Feb 2019
  65. Pehrsson E (2018) The Meaning of the GDPR Article 22. *Eur Union Law Work Pap* 31:37
  66. (1985) Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data
  67. (2017) *Big Data, Artificial Intelligence, Machine Learning, and Data Protection*. Information Commissioner’s Office

68. Jacobson PD Medical Liability and the Culture of Technology. Pew Charitable Trusts
69. Vladeck DC (2014) Machines without principals: Liability rules and artificial intelligence. *Wash Law Rev* 89:117–150
70. Parasuraman R, Riley V (1997) Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum Factors* 39:230–253. <https://doi.org/10.1518/001872097778543886>
71. Lyell D, Coiera E (2017) Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc* 24:423–431. <https://doi.org/10.1093/jamia/ocw105>
72. Goddard K, Roudsari A, Wyatt JC (2012) Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc JAMIA* 19:121–127. <https://doi.org/10.1136/amiajnl-2011-000089>
73. Fenton JJ, Taplin SH, Carney PA, et al (2007) Influence of Computer-Aided Detection on Performance of Screening Mammography. *N Engl J Med* 356:1399–1409. <https://doi.org/10.1056/NEJMoa066099>
74. Lehman CD, Wellman RD, Buist DSM, et al (2015) Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern Med* 175:1828–1837. <https://doi.org/10.1001/jamainternmed.2015.5231>
75. Winfield AF, Jirotko M (2018) Ethical governance is essential to building trust in robotics and AI systems. *Philos Trans Math Phys Eng Sci* 376:
76. Morning Consult (2017) National Tracking Poll 170401. Morning Consult
77. Bonnefon J-F, Shariff A, Rahwan I (2016) The social dilemma of autonomous vehicles | *Science*. *Science* 352:1573–1576. <https://doi.org/10.1126/science.aaf2654>
78. Awad E, Dsouza S, Kim R, et al (2018) The Moral Machine experiment. *Nature* 1. <https://doi.org/10.1038/s41586-018-0637-6>
79. Agar N (2004) *Liberal Eugenics: In Defence of Human Enhancement*, 1 edition. Wiley-Blackwell, Malden, MA 1242
80. Steinbrook R (2009) Controlling Conflict of Interest — Proposals from the Institute of Medicine. *N Engl J Med* 360:2160–2163. <https://doi.org/10.1056/NEJMp0810200>
81. Institute of Medicine (US) Committee on Conflict of Interest in Medical Research, Education, and Practice (2009) *Conflict of Interest in Medical Research, Education, and Practice*. National Academies Press (US), Washington (DC)
82. (2018) *Protecting Patients, Preserving Integrity, Advancing Health: Accelerating the Implementation of COI Policies in Human Subjects Research*. American Association of Medical Colleges
83. Bero L (2017) Addressing Bias and Conflict of Interest Among Biomedical Researchers. *JAMA* 317:1723–1724. <https://doi.org/10.1001/jama.2017.3854>
84. Fineberg HV (2017) Conflict of Interest: Why Does It Matter? *JAMA* 317:1717–1718. <https://doi.org/10.1001/jama.2017.1869>