

智能代理是刑事程序中的证人吗？

[瑞士] 萨比娜·格雷丝 / [德国] 托马斯·魏根特^{**} 著 樊文^{***} 译

【摘要】人工智能（AI）驱动的“智能设备”日益成为刑事程序所用证据的信息源。日趋复杂的机器学习方法产生了越来越多的人工智能生成的证据，这种自主生成的数据（比如使用现代汽车时生成的数据）的重要性可能会变得愈发突出。本文研究了来自人工智能系多个应用领域的此类证据的可采性，解释了如何将此类系统生成的数据作为证据引入刑事程序，并在讨论可能的解决方案之前，特别是讨论人工智能系统成为自身的证人并接受其他人工智能系统可靠性审查的可能性之前，分析了这类数据在诉讼法上的相关证据问题。

【关键词】人工智能系统；刑事程序中的证据；刑事程序中证据的可采纳性；对质权；刑事程序中的证人；刑事程序中的专家证人；不自证其罪的权利；刑事程序中作为证据的数据

【中图分类号】 D918.9

【文献标识码】 A

【文章编号】 1674-1226 (2024) 06-0559-15

Intelligent agents as a sort of witness in criminal proceedings?

【Abstract】'Smart devices' empowered by artificial intelligence (AI) are increasingly the origin of information used as evidence in criminal proceedings. The importance of such autonomously generated data, for instance data generated when using modern automobiles, is likely to become even more important as increasingly complex methods of machine learning lead to more AI-generated evidence. This article reviews the admissibility of such evidence stemming from various areas of application of AI-systems. It explains how data generated by such systems could be introduced as evidence into criminal proceedings and, analyses the relevant evidentiary problems, before it discusses possible solutions, in particular the possibility of AI-systems becoming a sort of witness of its own and being vetted for reliability by other AI-systems.

【Key Words】 AI-systems; Evidence in criminal proceedings; Admissibility of evidence in criminal proceedings; Right to Confrontation, Witnesses in criminal proceedings; Expert witnesses in criminal proceedings; Privilege against self-incrimination; Data as evidence in criminal proceedings

*“Intelligente Agenten als Zeugen im Strafverfahren?” 首发于 JZ 12/2021 S. 612-620. 本文的翻译和发表取得了作者和出版社的授权。

** [瑞] Prof. Sabine Gless, 瑞士巴塞尔大学法学院刑法学和刑事诉讼法学教席教授德国马普犯罪、安全与法律研究所“刑事司法中的算法解析和自动决策”项目小组高级研究员; [德] Prof. Thomas Weigend, 德国科隆大学法学院刑法学和刑事诉讼法学荣休教授。

*** Prof. Wen Fan, 德国弗莱堡大学法学博士, 中国社会科学院大学法学院副教授, 中国社会科学院法学研究所副研究员。

智能代理（Intelligente Agenten, IA），如机动车中的车道偏离警告系统，并不适合刑事程序中的传统证据手段的检验图式，将其观察结果用于证据的目的会引起严重的法律问题。本文介绍智能代理的各种相关应用领域，并解释以何种形式能够将其“感知”引入刑事程序。在对证据法上的问题进行分析后，文章讨论可能的解决方案。

一、由于智能代理而认定有罪

2016年，瑞士媒体报道了一位政治家因其车道偏离警示系统的不利“陈述”而受到严重过失伤害罪的刑罚处罚。该被告人受到的指控是，在过度疲劳的状态下驾驶车辆，与一名骑自行车的女性相撞。据说驾驶员辅助系统在事故发生前曾多次警告过该政治家与疲劳有关的驾驶错误¹。被告人最初否认这一指控，但后来接受了刑事处罚令，并指出睡眠呼吸暂停症*是（事故）发生的一个可能原因，而在此之前他并不知道自己有这一问题²。驾驶员辅助系统的“陈述证词”是否有助于在人类认知能力失灵的地方实现正义？还是说驾驶员就此自己认了，因为他无论如何也无法针对车道偏离警示系统*提供的证据成功地进行自我辩护。

今后，其他一些道路交通参与者也可能会遇到这位（瑞士政治家）驾驶员的情形。因为智能代理（IA）的使用正在增多，特别是在道路交通领域。例如，从2022年起，在欧盟注册的新车必须配备各种驾驶员辅助系统，如车道偏离警示系统和睡意警报，以提高道路交通安全³。这可能仅仅是个开始——我们越是生活在更多诸如机器人和隐形机器人陪伴的社会，随着它们不停地感知和处理其周围环境中的事件，它们被用作刑事程序中事后查明手段的潜力就越大。在人与机交互界面上记录行为的智能代理，为测量和评估提供了许多可能性——不仅是为了提高驾驶安全，也是为了有效预防⁴和追诉犯罪行为。

二、智能代理的运作方式和应用领域

车道偏离警示系统和睡意警示系统是驾驶员辅助系统，即嵌入机动车中的IT系统，它通过独立收集、处理数据、以算法评估数据，最终在此自主领会的基础上采取行动，在特定的驾驶情形为人类驾驶员提供帮助和支持。今天，该系统大多是这样设计的：通过机器学习来准备相关数据的算法处理——传感器学习如此“看”。为应用于道路，基于规则对传感器功能进行编程，以便在置入

¹ <https://www.watson.ch/schweiz/gesellschaft%20&%20politik/627658357-ex-fdp-chef-philipp-mueller-wegen-fahrlaessiger-schwerer-koerperverletzung-verurteilt>; <https://bi.ly/3wsPqVM>.

* Schlafapnoe, Sleep Apnea, 睡眠呼吸暂停，又称睡眠呼吸中止症。其症状的主要病征表现为，日间专注力、认知能力减低，白天犯困瞌睡，甚至强迫入睡。

² <https://www.berneroberlaender.ch/news/standard/nach-autounfall-philipp-mueller-verurteilt/story/10546602>; <https://bit.ly/3oE3CJ5>.

* Lane Departure Warning System, LDWS, 是一项汽车驾驶安全辅助系统。当传感组件检测到车辆偏离车道时，若驾驶员因精神不清或者疏忽而未打下转换车道的方向信号灯，系统会发出警示信号以提醒驾驶员返回车道。——译者。

³ Verordnung (EU) 2019/2144 v. 27.11.2019, ABl. Nr. L 325, 16.12.2019, S.1.

⁴ 如果考虑到未来，比如“暴力攻击倾向侦测系统”也是可以想象的，它可以根据某些参考点（驾驶员的鲜明个性、交通密度、速度、一天中的时间、音乐偏好）确定驾驶员的暴力攻击潜质，参见 *Hegde/Rokseth Safety Science* 122 (2020), Nr. 104 492.

车辆之前把系统的深入学习能力“锁定”在正常情况之下⁵。这样，程序员和制造商就不仅能确定，如何为其预期目的优化设计和训练驾驶员辅助系统，而且还能确定以机器学习为条件的暗箱（black box：黑匣子）究竟有多大，以及在虚拟通道中和通过训练测试的功能在何种条件下足以上路，或者说还需要进一步优化⁶。

复杂的辅助系统只有借助人工智能*（Künstliche Intelligenz, KI）⁷和强化的独立学习，才能以适应驾驶员的最佳方式自主地应对不同的交通状况⁸。即使这些系统事先（在虚拟通道和试车期间）通过监督下学习的方式进行了训练，例如识别某些交通标志或测量与车道标线保持规定的距离，也不能百分之百保证它们以后在所有交通情况下都不会产生错误的后果（结论）。因此，理想的情形或许是：这些系统在运行过程中能够进行自我优化，也可以是在人类驾驶员的介入下进行优化。此外，考虑到不同的目标群体，汽车制造商也可能感兴趣的是对驾驶员辅助系统——在规定的安全范围内——进行个别化调适，例如，为规避风险和更激进的女司机⁹。不过，这或许与机动车的批准车型有些不一致，特别是由于人类通常无法预见或者理解：基于人工智能的驾驶员辅助系统具体是如何开展其学习过程的。对于驾驶员，这也可能是个问题，对于他们来说，特定的行动——如对疲劳发出警告或者在特定驾驶情形下进行干预——来自至少不明其内在究竟的暗箱。

2018年在亚利桑那州发生了可能是第一起涉及自动化驾驶车辆撞死一女性行人的事故，该事故的报道直观地说明了智能代理作出决策的不透明性和由此对法规所产生的问题：尽管有许多技术上固定下来的安全规范，但事后无法确定为什么车辆的软件将一位推着自行车的女性行人归为“不明物体”¹⁰。这起事故给该车的制造商和运营商并没有带来任何刑法上的后果；而车的女测试员则

⁵ Reif et al., 《驾驶稳定系统和驾驶辅助系统（Fahrstabilisierungssysteme und Fahrerassistenzsysteme）》，2010, 206-209; Maurer et al., 《自动化驾驶（Autonomes Fahren）》，2015, 202; Winner et al., 《驾驶员辅助系统手册（Handbuch Fahrerassistenzsysteme）》，2012, 17-19, 45.）

⁶ 关于虚拟通道中的车辆辅助系统的培训，参见 www.economist.com/technology-quarterly/2020/06/11/driverless-cars-show-the-limits-of-todays-ai.

* 关于 IA 和 AI 的区别，作者的理解是：人工智能（AI）是任何形式的不是由人执行的独立自主地数据处理；而智能代理（IA）是借助于人工智能运作的对象或者数据集（Datensammlungen）。举例来说：疲劳警报器或者警告装置（Müdigkeitwarner）是一个智能代理，它利用其内置的人工智能独立记录、录制、处理数据并从中产生结果。——译者。

⁷ 关于定义，参见 Steege SVR 2021, 2 f.; Hilgendorf, in: Fischer (Hrsg.), Beweis, 2019, 229, 232. Art. 3 (1) (1) des Entwurfs einer Regulation on a European Approach for Artificial Intelligence [21.4.2021 COM(2021) 206 final 2021/0106] 包含如下定义：‘artificial intelligence system or AI system’ means software that (...) can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy. (...) [“智能代理系统或 AI 系统”是指（……）能够针对一组特定的人类定义的目标，产生诸如影响现实或虚拟环境的内容、预测、建议或决定的软件。智能代理系统被设计为不同程度的自主性运行。（……）”]。

⁸ 比如，戴姆勒公司的注意力辅助系统，就是个睡意检测系统，它向驾驶员发出警告，防止他们打盹。参见 Daimler (2021), ‘Attention assist: Drowsiness-detection system warns drivers to prevent them falling asleep momentarily’. <https://media.daimler.com/marsMediaSite/en/instance/ko/ATTENTION-ASSIST-Drowsiness-detection-system-warns-drivers-to-prevent-them-falling-asleep-momentarily.xhtml?oid=9361586>.

⁹ 例如，戴姆勒公司提供了根据用户的愿望和国家的习俗来调整车道偏离警示系统的敏感性 <[moba.i.daimler.com/baix/cars/177.0_mbox-high_2018_a/en_GB/page/ID_c198e8b2deb49d14354ae3655a22d369-68add78aec700212354ae3652093af55-en-GB.html](https://www.daimler.com/baix/cars/177.0_mbox-high_2018_a/en_GB/page/ID_c198e8b2deb49d14354ae3655a22d369-68add78aec700212354ae3652093af55-en-GB.html)>; <https://www.spiegel.de/auto/fahrberichte/kia-sorento-im-test-dieses-suv-laesst-sich-fernsteuern-a-0f1c1ea0-eb52-4bd0-95eb-c5f37bab055f>; <https://www.consumerreports.org/car-safety/lane-departure-warning-lane-keeping-assist-guide/>.

¹⁰ 高速公路事故报告 NTSB/HAR-19/03PB2019-101402 on Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018, 16, 可访问网址：www.nts.gov/investigations/AccidentReports/Reports/HAR1903.pdf.

在 2021 年受到审判¹¹。

如果想出于证据目的使用智能代理，就必须考虑到它们的运作方式与人类根本不同：驾驶员辅助系统的传感器不能像人眼一样识别对象。2016 年的一个训练过程典型地反映了人类的识别和机器识别之间的差距：训练一个智能代理来区分照片中的雪橇犬和狼。训练中使用的所有狼的照片都以雪作为背景。结果该智能代理将“雪”记录成了关键的区分特征，在后来的分辨过程中也就将所有带有雪的哈士奇照片归入了“狼”的类别¹²。“符合逻辑，但不合理。这难道不就是机器人的定义吗？”¹³ 艾萨克·阿西莫夫 (Isaac Asimov) 早在 1957 年就这么写了。

分类不正确会导致道路交通中的许多问题，例如，当传感器系统检测到一个“停车、让行！(Halt. Vorfahrt gewähren!)”字样的道路标牌时，它在机器训练中还能正确识别的，在真实行驶中却把它当作了“先行 (Vorfahrt)”标牌，而坐在车里的人却无法察觉到它的这个错误¹⁴。而且，即便驾驶员怀疑她的车在识别交通标志方面有问题，她也无法与智能代理交流，告知她自己的怀疑，因为智能代理不能解释清楚它们自己作出的决策，也无法反映或者呈现它们被预先编程的隧道视野 (Tunnelblick)¹⁵。

让人类能够理解智能代理决策过程的一种方法是自动化驾驶数据存储系统 (DSSAD*)，该系统以预定的时间间隔记录特定数据，以便以后能够检索这些数据，并且能够以此追踪智能代理的观察、评估判断和决定。借助于自动化驾驶数据存储系统，程序员和制造商就为他们的智能代理在某种意义上创设了一个“大脑”，以此赋予它们感知能力和记忆能力。至于之后驾驶员辅助系统“记住”了什么，取决于在道路上行驶时自动化驾驶数据存储系统决定以什么样的频率记录、处理和存储哪些数据。

三、智能代理的观察是刑事程序中的证据种类 (Beweismittel) 吗？

不断观察 (“目击”) 周围环境并存储收集信息的智能代理在澄清刑事程序中的事实方面可以发挥巨大的价值¹⁶。智能代理比人类证人更加可靠，因为他们不会被同乘 (Mitfahrer) 分心，不会疲倦，不会忘记任何事情，也没有可能为了追求自我利益而伪造所感知到的信息。因此，在审问式的刑事程序中，考虑到法院的全面之事实查明义务 (《刑事诉讼法》第 244 条第 2 款)，此类证据乍看起来似乎都是必须要获取的。但与此同时，出现的是多种法律问题：一方面是关于这种——设计时着眼于用户和制造商需求的——智能代理作为刑事案件证据提出的可靠性，另一方面是关于程序

¹¹ 关于指控，参见 <https://www.maricopacountyattorney.org/CivicAlerts.aspx?AID=751>。

¹² Ribeiro/Singh/Guestrin, “Why Should I Trust You?”, Computer Science (2016), S. 8. <https://arxiv.org/pdf/1602.04938.pdf>。

¹³ Asimov, The Naked Sun, 1957, 232.

¹⁴ 比方说，对道路交通标识的操纵，欺骗了传感器，但人眼是看不出来的，参见 Eykholt et al., Physical Adversarial Examples for Object Detectors, Computer Science (2018) <https://arxiv.org/pdf/1807.07769.pdf>。

¹⁵ 对此参见 Eitzkorn MMR 2020, 360, 361; Hilgendorf, in: Fischer (Fn. 7), 233 f.

“隧道视野”，是个心理学概念。一个人如果处于隧道之中，视野自然被限制在狭窄的区域之内，意指受限于个人经验和见识，无法看到更为广阔的世界，永远无法做出超出见识之外的事情。——译者。

* Data Storage System for Automated Driving. ——译者。

¹⁶ 这不仅适用于驾驶员辅助系统。多年来，美国的媒体一直在报道关于使用健身手表 (<https://www.wired.com/story/health-fitness-data-privacy/>) 和智能心脏起搏器 (<https://arstechnica.com/science/2017/02/ohio-mans-pacemaker-data-betrays-him-in-arson-insurance-fraud-case/>; <https://www.wired.com/story/your-own-pacemaker-can-now-testify-against-you-in-court/>; <https://www.bbc.com/news/technology-40592520/>) 数据的评价争议。这些对象可以通过运动特征和身体活动数据证明嫌疑人有罪。

的公正性。而现行德国刑事诉讼法将严格的证明程序*中的证据种类限制于证人、鉴定人（专家证人）、文书和勘验¹⁷。这就出现了一个问题：即智能代理的观察是否可以归入其中的一类，如果不能，它们的感知是否可以通过其他方式引入刑事程序。

（一）是证人吗？

正如开头所述的案例，如果来自驾驶员辅助系统的疲劳警告记录被用作驾驶员不能驾驶的证物，就必须考虑智能代理是否能作为证人的一种（出庭作证），因为他们能够通过再现自己对过去事实和状态的感知来帮助发现真相¹⁸。但是，现行诉讼法显然认为，只有自然人才可以成为证人；从诸如关于询问种类、对于个体人的细节规定和宣誓义务的规定中都可以看出这一点（《刑事诉讼法》第 57、59、68、68a 条）。因此，尽管德国的俗语把监测交通的雷达装置可能称为“铁皮警察（Blechkopisten）”，但它们仍然不能作为证人出庭。车上的雷达装置和车道保持助手都不能在法庭上做言词陈述，也不能回答法官和其他诉讼参与方的提问。这只有在（未来）创造出不仅能沟通而且能反映其自身数据生成的智能代理时，才有可能。虽然很多研究活动都是针对“可解释的智能代理”，但迄今为止的进展还不足以让人们期待，具有语言功能的机器人在可预见的未来能够作为证人进行陈述。

（二）是勘验证据吗？

即使驾驶员辅助系统由于没有沟通能力而不能被视为证人，智能代理产生的记录也有可能被评价为勘验客体（Augenscheinsobjekte），类似于测速雷达证明超速事实的“抓拍照片”。在一开始讨论的案例中，可以考虑将数据可视化，借助这些数据，车道偏离警示系统的反复警告可以作为视觉文件（信号警示！）呈现，法院可以在主审程序中与程序参与各方进行讨论¹⁹。

从今天的角度来看，只要驾驶员辅助系统是一个基于规则的系统，其数据处理是预先编程好的，就是可以理解的，这在技术上是完全可以想象的。第一个基本前提是 DSSAD 的数据生成和存储的标准化，类似于雷达设备领域的标准化²⁰。但是，当一个智能代理——如睡意警示系统——生成反映其对情形评估的评价性数据时，过程的可视化就变得很难。例如，综合车道偏离和睡意警告系统可以根据记录的测量数据得出评估，即汽车无法可靠地在车道上直行，人类驾驶员在驾驶座上（坐姿）不再完全直挺，以及他或者她的眼睛多次打盹。因此，一个脉冲便被自动传送到车轮，使车进入被车道偏离警告系统确定为“正确”的车道，并向驾驶员发出警告。这个过程与原始数据和测量数据（可能还有评估数据）一起被记录在 DSSAD 中，正如智能代理通过声音和 / 或视觉信号向人类驾驶员发出的警告。不过，智能代理从观察到“自己”行动的这个过渡，事实上是否能真正直观地呈现出来，是值得怀疑的²¹。为此，必须有可能展示并说明智能代理在哪些点上不仅进行了测量，而且

* 与严格的证明程序（Strengbeweisverfahren）并列的是自由证明程序（Freibeweis-Verfahren）。严格的证明程序是指在主审程序中，适用于所有实体法上重要的事实的证明程序。这就是说，严格控制：可以（允许）使用哪些证据材料（Beweismittel），即仅仅是证人、鉴定人、文件和勘验。此外，严格规范的是证据采信的程序（Beweisaufnahme），即怎么对待这些证据种类的问题。而自由证明程序，既没有证据种类排除的基本原则，也没有其运用的规定规则。它适用于侦查程序，主审程序中仅仅程序上重要的事实（比如，被告人的受审能力）以及上诉程序。——译者。

¹⁷ Gössel, in: Löwe/Rosenberg, StPO, Band 1, 27. Aufl. 2016, Einl. L Rn. 3.

¹⁸ 关于证人的概念参见 Kühne, 《刑事诉讼法（Strafprozessrecht）》，9. Aufl. 2015, § 53 Rn. 792ff.; Eisenberg, 《刑事诉讼的证据法（Beweisrecht der stop）》，10. Aufl. 2017, Rn. 1000 ff.

¹⁹ Vgl. Huckenbeck/Krumm NZV 2017, 453.

²⁰ 关于欧盟经济委员会这个方面的谈判，参见 ECE/TRANS/2021/14.

²¹ 本文注脚 10 所提到的来自美国的调查报告清楚地说明，智能代理的“决定”只能得到有限的溯源。

进行了评价, 以及是基于什么进行的评价。

如果是人工智能基础上的系统出现这种情形, 比如在可预见的将来可期待的系统, 情况还要更加复杂²²。这种驾驶员辅助系统收集各种原始数据作为参考变量, 比如汽车与车道标线的距离, 并借助给定的算法对这些数据进行程序化处理。它的源代码能使智能代理优化自身以完成任务, 这样它就能够根据人类无法查验的一个过程作出具体个别的情景反应。似乎很难想象对这些复杂的决策过程进行标准化的视觉表达, 因此, 在这方面几乎不可能提供勘验证据*。

(三) 是鉴定者证据吗?

对于事故后汽车中的数据痕迹的评估, 鉴定者证据(Sachverständigenbeweis)在许多方面是有帮助的, 并且越来越成为常规。与勘验证据的情况不同, 法官在这里面对的是一个人, 他在一定程度上负有解释驾驶员辅助系统生成的数据的责任, 并且可以回答与案件的具体事实有关的询问(Rückfragen), 通过与模拟世界的相似性使法官能够理解。此外, 辩护方不仅可以对检察院或法院指定的鉴定人(专家)进行批判性质疑, 还可以根据《刑事诉讼法》第244条第4款申请提名指定另外的鉴定人(专家), 或者提供自己的鉴定人(专家)。这样就提高了准确性之保证, 并为诉讼程序引入一个强有力的辩护因素。

上述就主张了, 通过鉴定人(专家)鉴定的方式也要把智能代理的观察和结论引入诉讼程序。这里专家鉴定就包括关于系统运作的信息, 用于法庭陈述的数据的有效性, 特别是关于它们的生成、储存和可复制性的信息²³。此外, 专家的专业意见书还必须涉及这些问题: 如何理解智能代理生成的数据, 以及他们以多大的可能性实现了对事实评估重要的选择性假设(例如: “人类驾驶员太累了, 无法正常驾驶” vs. “驾驶员辅助系统没有正常运作”)。不过, 对于这一领域的鉴定者证据来说, 一个必要的前提是, 法律人(律师)和法庭科学家之间要有一套共同的术语, 以便能够对智能代理生成的数据进行归类, 且能充分考虑到不同数据的特点, 比如测量数据和“评价数据”。

即便如此, 专家们立足科学的专业知识对从驾驶员辅助系统中获取的数据之“可信性(Glaubhaftigkeit)”问题能够在多大程度上发表意见, 仍然是值得怀疑的, 而这个问题对事实重构至关重要。在一个基于规则的系统, 对数据进行可理解的验证(Validierung)似乎是可能的。专家们可以通过解释系统的工作原理和各种数据类别的有效性来使他们为法庭陈述所做的评估变得可

²² 参见 *Wisselmann*, 《自动汽车的法律侧面 (Rechtliche Aspekte automatisierter Fahrzeuge)》, in: Hilgendorf/Hötitzsch/Lutz (Hrsg.) Beiträge zur 2. Würzburger Tagung zum Technikrecht im Oktober 2014 (2014年10月维尔兹堡第二届技术法会议文集), 2015, 11; Nolting, 《汽车工业中的人工智能 (Künstliche Intelligenz in der Automobilindustrie)》, 2021, 8 ff., 121 ff.; European Automobile Manufacturers Association, ACEA, Position Paper Artificial Intelligence in the automobile industry, November 2020 <<https://www.acea.be/publications/article/position-paper-artificial-intelligence-in-the-automobile-industry>>. 当前的改革考虑, 参见 *Hilgendorf* JZ 2021, 444, 449f.

* Augenschein (real evidence) 的客体是影印、复印, 电影, 视频/唱片和录音带/笔记、记号/文件/人/流程和实验(行驶试验, 事件过程的重建)。用 Augenschein 进行证明, 只能由法官进行, 可以在主审程序内, 也可以在主审程序外。其他人如果进行了查看(比如检察官、警察), 他们只能是以证人进行。不同于证人证据(第250条), Augenscheinsbeweis 不必直接展示。这会有两个后果: 1、Augenschein 的这种证据形式, 可以被其他证据形式代替, 比如, 法官并不查看犯罪地点, 而是查看地点的照片; 不播放录音带, 而是宣读其内容的记录; 可替换的界限, 由查明义务划定。2、法官可以把 Augenschein 的工作委托给 Augenschein 的帮助者。如果事实或者法上的理由都不妨碍法官进行查验, 法官也可以自己进行查验, 或者法官认为应该得到协助, 可以委托帮助者。比如, 查看脚手架顶端情况的建筑工人; 查看沉船残骸的潜水员。鉴定人也可以受托这些工作, 但是他们报告的是他们作为证人的感知。——译者。

²³ *Vuille/Arnold*, Strassenverkehr/Circulation routière 2015, 51, 53 f.

信。不过，鉴定者证据可能更为困难的情况是在于以人工智能（AI）为基础的“自主”系统，因为在这里，由于暗箱的问题（Black-box-Problematik），专家不能完全回溯呈现并让人理解通向某种评估（“驾驶员疲劳”）的路径和智能代理（IA）的决策（车道偏离警示系统的干预）。此外，如果DSSAD不能充分存储相关数据，或者这些数据不能被可靠地读取出来，鉴定者证据就达到了极限。例如，如果一个综合的车道偏离警示系统和睡意警示系统把无法标准化测量的参考点（如驾驶员的身体紧张状态和眼睛的眼睑运动）纳入进来，然后独自评估它们而并不清楚为什么这样做，就可能是这种情况。除此之外，如果智能代理制造商（IA-Hersteller）以商业秘密为由不公开DSSAD数据，那就可能出现重大的证据漏洞²⁴。如果证据法上要求测量链条和判断链条的全面可（查核）验证性，那么（智能代理）作为一种鉴定者证据就只能在特定条件下来考虑。

（四）是特殊的（sui generis）证据种类吗？

正如我们所看到的，智能代理与证人有相似之处，但它们也与人类应答者有重大的不同。虽然它们能传达一些关于它们感应到的东西，但它们还不能像人类证人那样被询问，所以无法检验核实它们陈述的可信性。在这个方面，它们更像是今天用于监控遵守道路交通法规的机器。但是，智能代理也不同于雷达设备之类，它不是为人类可理解的特定测量过程而设计的，而是根据大量不同的数据独立地决定，何种条件下警告驾驶员注意疲劳（驾驶）。只要智能代理能够合规地自主完成其任务，这种能力的代价将是暗箱问题，这个问题原则上会阻碍其在刑事程序中被用于重构案件事实：智能代理不能解释他们的观察和决策；他们也不能批判性地反思他们的评估，因此不能提供任何关于可能存在的误解或者关于其决策过程的信息。2021年《欧盟人工智能条例草案》由此曾设定了这样的原则：在没有人工审查可能性（human oversight：人类失察）的情况下不得使用“高风险人工智能系统”的陈述，而且必须始终保持人类纠正或者忽略智能代理输出结果的可能性²⁵。

鉴于法院的任务是查明真相，只要智能代理的观察可以被认为是可靠的，而且诉讼程序的公正性继续得到保证，那么利用智能代理的观察来澄清案件事实就是有好处的。智能代理就可以成为新一代证据种类的一部分。迄今为止，德国《刑事诉讼法》中还没有规定这种特殊的（sui generis：拉丁语，“自成一类”，意指独特性—译者注）证据种类，这个形式问题可以通过修法而相对容易地予以解决。但是，更为重要或者说更具分量的是向“智能代理”的证据种类可能提出的实质问题。

四、诉讼法上的问题

无论智能代理在刑事诉讼证据种类目录中如何归类，在刑事程序中使用智能代理将会出现四个主要的程序法问题：由于借助智能代理的证据违背被告人他自己的意愿，是否使得被告人变成对自己不利的证据手段？智能代理是否侵犯了个人隐私的核心区域？被告人能否充分审查智能代理“展示的证据”并在法庭上提出质疑？就其本质而言，智能代理是不是一种支持理性判决的充分可靠的证据种类（Beweismittel）？

（一）智能代理提出证据情况下的“不自证其罪（Nemo tenetur）”

任何人都没有义务对自己进行控告（nemo tenetur se ipsum accusare）原则与人的尊严和公正审

²⁴ 参见 *Schlanstein*, NZV 2016, 201, 204; *Roth*, Yale Law Journal 126 (2017), 1972, 1987ff.; *Wexler* Stanford Law Review 70 (2018), 1343;

²⁵ 参见 Art. 14 Abs.1, Abs.4 des Entwurfs einer EU Regulation laying down harmonized rules on Artificial Intelligence (Fn. 7).

判原则(Fair-trial-Grundsatz)有关,它保护个人在刑事程序中不必违背自己的意愿积极主动地置自己于不利(认罪)²⁶。例如当车道偏离警示系统观察人类驾驶员并拍录他们的行为方式及他们对辅助提醒的反应时,乍一看并不触及该原则。毕竟,类似的信息在今天已经受到评价,比如,在制动操作之后,从汽车轮胎的磨损中可以看出的刹车痕迹,或者在其他领域,由传统的像机或录音带固定的被告人的行为或者言语,就以勘验证据或者鉴定者证据的形式被引入刑事程序。人们并没有把这视为违反不自证其罪原则(Nemo-tenetur-Grundsatz)²⁷。根据主流观点²⁸,这甚至适用于秘密制作的磁带录音,因为这里并不要求当事人自己积极地置自己于不利(入罪),而只是对自愿实施行为的录音的评价,这自愿实施的行为可能成为可罚举止(入罪)之不利证据²⁹。

当然,在不明晰的中间区域有国家强制规定的观察机制,比如,根据《(欧盟)2019/2144号条例[VO(EU)2019/2144]》,从2022年起所有新注册的汽车都要安装驾驶员辅助系统³⁰。既然驾驶人员在此无法合法地逃避持续的(系统)观察,而且也明知这一点,完全可以说他是受到了国家的强迫,要交出(即放弃或披露包括)驾车期间于己不利(自证其罪)的信息。而对此仅仅以下这点就能反驳:不是任何人都必须驾驶汽车,而是还可以使用公共交通或者选择步行³¹。

(二) 侵入人格核心区域

智能代理正在不断切近人类。一想到出于医疗需要而植入人体的智能代理,就有比如“智能”心脏起搏器、胰岛素泵或者假肢。借助外部可操作的仪器,甚至可以读取和评价脑电波;在这方面已经在考虑“数字化测谎仪”的应用³²。从宪法的角度来看,这里的问题是,智能代理会侵入私人生活的核心区域,并收集属于个人最内在的私密领域的信息。众所周知,立法机构和联邦宪法法院严禁在国家刑事程序中使用这种信息,因为这涉及到侵犯人的尊严³³。在这个方面,对于智能代理生成的信息在证据法上的可应用性存在着边界限制。

由于智能代理融入我们的日常生活,未来可能会产生许多新的边界问题(Grenzfrage)³⁴。例如,在过去,智能手机是日用品,对它的搜查从人格核心区域的角度来看似乎不成问题。因此,被置入比如被告人的智能手机或者汽车中的辅助系统,虽然人们也在电信监控方面对它进行法律分析(《基本法》第10条,《刑事诉讼法》第100a及以下条款),但是,原则上不会被判定为侵犯人格范围的核心区域³⁵。但如今智能手机不仅被许多人一直随身携带,而且还包含——类似于一些智能手表——非常私密的数据,国外已经有各种声音要求对(目前)这种法律状况进行新的评估,以便考

²⁶ BVerfG NJW 2013, 1058; Gless, in: Löwe/Rosenberg-StPO, Band 4, 27. Aufl. 2019, § 136 Rn. 27; 详细的论述和批判,参见 Rösinger,《被告人免于不利于自我的强制的自由(Die Freiheit des Beschuldigten vom Zwang zur Selbstbelastung)》,2019,48 ff.

²⁷ 参考 Arnold, Digitale Spuren im Strassenverkehr — die Zukunft hat begonnen!, in: Schaffhauser (Hrsg.),《道路交通安全法年鉴(Jahrbuch zum Strassenverkehrsrecht)》2017, S.381 ff.

²⁸ Wolfslast, NStZ 1987, 103; Gless, in: Löwe/Rosenberg (Fn. 26), § 136 Rn. 27a, 35a.

²⁹ Vgl. Rösinger (Fn. 26), 237 ff.

³⁰ Gless StV 2018, 671, 672.

³¹ 对此,参见 Macula,《行政(监督)法上的协作义务和刑事诉讼上的免于自我不利的自由(Verwaltungs(aufsichts)rechtliche Mitwirkungspflichten und strafprozessuale Selbstbelastungsfreiheit)》,2016,6 ff.,74 ff.

³² 对此,参见 Gerhold, ZIS 2020, 431.

³³ Vgl. § 100d StPO; BVerfGE 109, 279; Gless StV 2018, 671 m.w.N.

³⁴ 对此,参见 Hilgendorf (Fn. 7), 238 ff.

³⁵ 参见 Hildebrandt/Koops, Modern Law Review 73 (2010), 428; Osula, International Journal of Law and Information Technology 24 (2016), 343; Lasagni, New Journal of European Criminal Law 9 (2018), 386; Lawrence William & Mary, Journal of Race, Gender, and Social Justice 26 (2019-2020), 77; Rottmeier/Eckel, NStZ 2020, 193, 197 ff.

虑推进这种法律状况的发展³⁶。从德国的角度来看，人们根本不可能把智能手机等同于身体的一部分，即使在一些人看来，由于不断地使用，它似乎已经长在了他们身上。即便智能手机或者智能手表持续不断地记录身体数据，它是否还能保留到对其评价之时，也确实是值得怀疑的³⁷。

（三）对智能代理生成的数据之对质权

把智能代理生成的、特别是评价性的数据用作刑事程序中的证据，在法治国家上的一个重大挑战是，来自于被告人在刑事程序中对不利于他的（指控）证据种类（Beweismittel）提出质疑的权利。

1. 《欧洲人权公约（EMRK）》第6条第3款d项

根据《欧洲人权公约》第6条第3款d项（the right to confront witnesses），每位被告人都有权“向对其不利的证人发问或者接受对方发问”。这一提法源于普通法地区（Common-law-Bereich）的对抗式刑事程序，其中有控方的证人和辩方的证人。而所谓的对质权已通过批准接受《欧洲人权公约》成为德国法律体系的一部分，并逐渐被以审问式为特色的德国刑事诉讼程序之实务判例所接受³⁸。

根据欧洲人权法院（EGMR）的实务见解，对质权意味着被告人可以向据称已经感知到相关事件的人进行发问，即不仅是在主审中原始证人传达给他的感知的传闻证人³⁹。根据其措辞，《欧洲人权公约》第6条第3款d项只适用于证人。然而，欧洲人权法院对“证人”一词的解释很宽泛，因此它也包括受伤者、鉴定专家“和其他在法庭作证的人”⁴⁰。欧洲人权法院从公平审判原则和相关的刑事程序中控辩双方武装对等的原则中，推导出了这样的要求：被告人必须有机会在法庭上询问控方指定的专家证人，也可以用自己的专家证人来对抗反驳控方证人⁴¹。在另一个判决中，法院更进一步认为，控方所依赖的重要证据（案例：据称是被告人伪造的支票）即使被（据称⁴²）以法院的名义（受法院之托）销毁，从而辩方无法获得（隐瞒辩方），也是对公平审判原则（Fair-trial-

³⁶ 参见 *U.S. Supreme Court, Riley v. California*, 573 U.S. 373, 381(2014): “Modern cell phones ... are now such a pervasive and insistent part of daily life that the proverbial visitor from Mars might conclude they were an important feature of human anatomy” (“现代手机……如今是日常生活中如此普遍而须臾不可离开的一部分，以至于传说中的火星来客可能会认为它们——现代手机是人类解剖学的一个重要特征”)。在英美法系中的讨论，参见 *Redmayne*, 27 *Oxford Journal of Legal Studies* 27 (2007), 209; *Bronshteyn*, *Stanford Law Review* 72 (2020), 455.

³⁷ 参考 *Blechschnitt*, MMR 2018, 361, 362 f.; *Hilgendorf* (Fn. 7), 238 ff.

³⁸ 对此，详见 *Weigend*, in: *Festschrift für Jürgen Wolter*, 2013, 1145, 1156 ff.

³⁹ 参见 *EGMR, Urteil v. 24.11.1986 – Unterpertinger ./. Austria*, Nr. 9120/80, Rn. 31 ff. 对 EMRK 第 6 条第 3 款第 d 项的实务见解的概述，参见 *Biral*, *European Journal of Crime, Criminal Law and Criminal Justice* 22 (2014), 331; *Decaigny*, *New Journal of European Criminal Law* 5 (2014), 149.

⁴⁰ *EGMR, Urteil v. 11.12.2008 – Mirilashvili ./. Russia*, Nr. 6293/04, Rn. 158.

⁴¹ 参见 *EGMR, Urteil v. 25.7.2013 - Khodorkovskiy and Lebedev ./. Russia*, Nr. 11082/06 und 13772/05, Rn. 711 ff.; *EGMR, Urteil v. 24.4.2014 – Ivanovski ./. FYR of Macedonia*, Nr. 10718/05, Rn. 57 ff. 对根据 EMRK 第 6 条鉴定证据的要求之详述，参见 *Vuille/Lupária/Taroni*, *Law, Probability and Risk* 16 (2017), 55.

⁴² 在同一案件的再审中，多年后出现了原始支票并提交给了法庭。欧洲人权法院驳回了被判刑人重新提出的申诉，理由是在新的诉讼程序中，他一直没有行使要求核实签名真实性的权利；*EGMR, Urteil v. 15.10.2009 – Papageorgiou ./. Greece no. 2*, Nr. 21032/08, Rn. 33-35.

Grundsatz) 的违反⁴³。

这一论理可以有保留地(cum grano salis)适用于智能代理。对罪责问题具有决定性意义的事实认定和结论,如果是由智能代理以独立的且部分无法核查的方式作出的,而且如果只给被告人机会去询问程序员或者鉴定专家——智能代理本身毕竟是为证明被告人罪责而提出的证据种类,那么,无论如何都不能说这种询问满足了对质权。不过,有待解决的问题是,被告人是否也能够从公平审判原则(Fair-trial-Prinzip)中推导出一项这样的请求权,即以可核查的方式向他披露智能代理的“决策”过程。

2. 对原始数据和源代码的权利(结合《基本法》第2条第1款与第20条第3款)。

这个问题——在复杂程度稍低的层面——是德国法院之间争议的对象,最近才由联邦宪法法院结合《基本法》第2条第1款和第20条第3款推导出的“原始数据权(Recht auf Rohdaten)”予以澄清⁴⁴。争议的焦点是,在违反交通条例程序中的当事人是否有程序法上的请求权,获得作为原始数据的雷达测量设备的记录以及该设备的算法处理所依据的源代码。针对在举证时要具体解释这些原始数据并将其提供给当事人准备其辩护的义务,个别法院认为,机关和法院有权相信经过校准和定期监测的测量设备的功能的可靠性,只要没有故障的迹象,就不必为重构个别案件的事实而审查这些设备⁴⁵。直到最近,大多数地方高等法院才拒绝了当事人要求告知原始测量数据和其他关于测量设备运作方式的材料的请求——理由基本上都是出于简化程序的考虑⁴⁶。

近年来,一些法院肯定了当事人的这种请求权,即他们有法上的听审权(受审权)*和获得公平审判权(《基本法》第2条第1款与第20条第3款、第103条第1款)⁴⁷。2019年,萨尔州宪法法院认为,如果雷达设备根本没有记录个人的测量数据,致使当事人也没有办法获得这些数据,则

⁴³ EGMR, Urteil v. 9.8.2003 – Papageorgiou ./ Greece, Nr. 59506/00, Rn. 33 ff. (Rn. 36: “The right to an adversarial trial means, in a criminal case, that both prosecution and defence must be given the opportunity to have knowledge of and comment on the observations filed and the evidence adduced by the other party. In addition, Article 6 § 1 requires that the prosecution authorities should disclose to the defence all material evidence in their possession for or against the accused.” “在刑事案件中,获得对抗性审判的权利意味着必须给予控方和辩方机会,了解和评论另一方提出的意见和证据。此外,第6条第1款规定,检察机关应向被告方披露其掌握的有利于或不利于被告的所有重要证据”)。也可参见 EGMR, Urteil v. 24.10.2007 – Baumeit ./ France, Nr. 56802/00, Rn. 52 ff. (如果检察机关随后在没有通知被告的情况下向法院转交文件,则违反了《欧洲人权公约》第6条第1款)。还可参见 EGMR, Urteil v. 6.5.2003 – Perna ./ Italy, Nr. 48898/99, Rn. 29 ff., 在这里,法院强调,被告的要求仅限于提供对案件裁判重要的证据种类。

⁴⁴ BVerfG NZV 2021, 41.

⁴⁵ BayObLG, Beschl. v. 12.01.1966 – RReg. 1b St 278/65; 进一步的文献证明,参见 Schrey/Haug, NJW 2010, 2917, 2918; Röß NZV 2018, 507.

⁴⁶ OLG Bamberg, DAR 2016, 337; OLG Oldenburg, NZV 2017, 392.

* 为保障必要的权利救济,公民能否以及在多大程度上能够诉诸司法,是一项公民权利或者人权;根据人权公约的要求,在裁判针对任何人的指控或者确定他在诉讼中的权利和义务时,每个人都有接受法庭审判的平等权利(平等受审权)。即是有合格的法庭而不是由其他任何机构或者实体来做出裁判,是必须得到保障的公民权利,而不是出自法院的任意选择,也不是直观地出自司法本身的“特性”。比如,《公民权利和政治权利国际公约》第14条第1款就明确规定:“法院和法庭面前人人平等。在裁定针对任何人的指控或者确定它在法律讼案中权利与义务时,人人皆有资格获得由根据法律设立的有权能的、独立、无偏倚的法庭所为的公正而公开的审判”。参见夏勇,宪法制度, in: 夏勇主编,《法理讲义》(下),北京大学出版社2000年版,第487页。——译者。

⁴⁷ VerfG Saarland NZV 2018, 275; KG, Beschl. v. 7.1.2013, 3 Ws (B) 596/12 – 162 Ss 178/12; OLG Celle NJOZ 2017, 559; LG Trier DAR 2017, 721. 由于地区高等法院在这个问题上的判例不同,莱茵兰-法尔茨州宪法法院认为,科布伦茨地区高等法院没有将这个问题提交给联邦最高法院(BGH),是对法定法官权利的侵犯; VerfG Rh-Pf NZV 2020, 92.

违反了公平审判的基本原则⁴⁸。诚然，标准化的测量程序在法律上具有类似于预期的专家鉴定意见的意义，专家鉴定意见通常为法官的判决提供可靠的事实基础；但这并不排除在具体个案中对其事实根据和科学结论进行的事后审查。有效辩护包括当事人可以查询是否存在以前不知道的、对行为指控的坚实性上的怀疑；否则，有罪判决的事实根据就被最终排除在法庭审查之外⁴⁹。在2020年的一个判决中，联邦宪法法院赞同这一评估，并从公平审判权中推导出当事人有权了解为调查目的而准备好的所有测量材料，即使这些材料没有被列入程序卷宗⁵⁰。这一判决起初没有改变这样一个事实，即法院在使用经过认证和校准的测量仪器的结果时，只需要说明所使用的测量程序（流程）和公差值（Toleranzwert），而不必涉及自己对测量方法可靠性所作的认定（结论）⁵¹。不过，新近的实务见解强调了数据处理过程和借助自动机器生成的测量结果的可追溯性理解（Nachvollziehbarkeit）的重要性⁵²。

现在，成问题的是，承认对测量程序细节的知情权，将会如何影响到智能代理生成的数据和评估的可使用性。人们可能会说，智能代理并不像测量仪器那样得到校准，因此，知道原始数据和源代码并不足以反驳由智能代理独立生成的评估。但是，这或许会使各当事人处于这样的境地：即使错误的评估近在眼前、显而易见，因为国家不能校准，也就不可指责他们的“控方证人”。显而易见的错误评估有可能特别就是如下这种情况：如果系统在机器学习的过程中“学习”评估诸如身体紧张、眼睑活动或“连续打哈欠”等参考点，而这些参考点在客观上是不完全可以追溯理解的并且易于被误解的，而智能代理的评估（如：“驾驶员过度疲劳”）是以这种参考点之间的复杂的相互作用为根据作出的。由于智能代理生成的人类行为评估的这些特殊性，出现了一个特别紧迫的问题，即如何使它们的使用与当事人或者被告人的宪法权利相协调，使他们能够通过获得完整的事实依据来了解可能的减免情节。如果将关于被告人知情权的最近实务见解严格适用于智能代理生成的证据，那么通常显然有理由否定其可使用性，因为被告人无法审查判决的事实根据。不过，也许可以开辟新的路径来弥补这一不足（见下文五）。

3. 智能代理的基础上形成法官的确信？

德国刑事程序法对有罪判决设定的前提条件是，法院以理性上有根据的方式已经确定了被指控者的罪责（《刑事诉讼法》第261条）⁵³。因此，法官必须能够为其对被指控者的罪责之确信提出客观的根据；仅仅是猜测或者推测是不够的⁵⁴。在此，法官也可以依靠科学所保障的假设，而这些假设的根据本身是无法全面核实的；但是，这些假设在各自的科学中必须被承认为基本上是没有疑问的⁵⁵。

如果（重要的）证据来自一个暗箱，而这个暗箱作出决定的标准和程序是诉讼各方和法官都

⁴⁸ *VerfG Saarland* NZV 2019, 414; 对此最早的评论，参见 *Gratz*, DAR 2019, 500; *abl. Bspr. König*, DAR 2020, 362.

⁴⁹ *VerfG Saarland* NZV 2019, 414. A.A. *OLG Köln* DAR 2019, 695, 696 („, 审判庭认为，如果允许有关人员在任何时候仅凭怀疑就质疑和审查，那么就会阻碍预期的专家鉴定意见的系统性，甚为荒唐)。

⁵⁰ *BVerfG* NZV 2021, 41.

⁵¹ 基础性论述参见 *BGH* NJW 1993, 3081; 也可参见 *OLG Hamm*, NStZ 1990, 546; *OLG Schleswig*, Urteil v. 11.11.2016 – 2 Ss OWi 161/16 (89/16); *OLG Köln* DAR 2019, 695.

⁵² *BVerfG* NZV 2021, 41. 以前的实务见解，参见 *VerfG Saarland* NZV 2018, 275; *OLG Düsseldorf*, Beschl. v. 20.1.2010 – 5 Ss (OWi) 206/09 (OWi) 178/09; *OLG Karlsruhe*, Beschl. v. 12.2.2010 – 3 (5) SsBs 629/09-AK 4/10; *KG* NZV 2010, 311; *OLG Karlsruhe* NZV 2020, 368; *AG Dillenburg* DAR 2009, 715; *AG Solingen* NJW-RR 2017, 866.

⁵³ 在英美法律中，有罪判决的合理性论证有这么一句话：必须“排除合理怀疑（beyond a reasonable doubt）”才能确定有罪。对此参见 *EGMR*, Urteil v. 2.1.2013 – *Virabyan* ./ Armenia, Nr. 40094/05, Rn. 150.。

⁵⁴ *BGH* NJW 1999, 1562; *BGH* NStZ-RR 2012, 256; *Miebach*, in: MüKo StPO, 2016, § 261 Rn. 61; *Ott*, in: KK StPO, 8. Aufl. 2019, § 261 Rn. 6, 62.

⁵⁵ *Miebach*, in: MüKo StPO, 2016, § 261 Rn. 70 f.

无法完全理解的，那么就不可能进行有充分理性及科学根据的证据证明。而在未来人工智能驱动的“自学的”智能代理中，出现的正是这个问题，即使在完全了解原始数据、测量数据和源代码中规定的算法处理的情况下，由于（至少部分）面向自学的过程的复杂性，人类也无法完全深入了解这种智能代理的评估和结论。再以那个疲劳警示系统为例子：事故发生后，虽然能够确定智能代理警告了人类驾驶员，但是，我们不可能具体了解这一警告是如何产生的，即辅助系统的评估是基于实际存在的疲劳迹象，基于对人身物理过程的错误解读，还是基于内部的错误推断（错误结论）。如果法院还是根据疲劳警示系统提供的信息来给驾驶员作有罪判决，最终就是以非理性的方式“盲目”地相信：智能代理正确记录了相关数据并从中得出了准确的结论，这可能会违反作出理性判决的基本原则。因此，疲劳警告虽然可以作为事件流程的间接证据（Indizien，推定证据），但仅其本身还支撑不起有罪宣告。

五、可能的解决方案

鉴于上述呈现的困难，刑事司法在努力查明案件事实的过程中，是否应该放弃智能代理生成的观察（证据）？或者说，这不是一种危险的法治国家纯粹主义（rechtsstaatlicher Purismus：有文献译为“法治国的洁癖”），尤其在交通犯罪中，这种证据往往可能比人类证人的陈述（证词）更为可靠，而放弃这种证据人们是否会陷入更大的错判风险？

（一）智能代理—证据（IA-Beweisen）可检验性的根本问题

正如所解释的那样，主要问题是如何验证智能代理的证词（Angaben）与——毕竟难以把握的——“真正的”现实是一致的。众所周知，对于证人，确定其陈述的可信性一直是一个大问题⁵⁶，但对证人至少还可以继续询问（质证询问），心理鉴定专家可以判断其陈述（证言）的内在一致性。如今对智能代理似乎不可能进行这样的查验：车道偏离警告系统或者睡意警告器不可能探究审查自己评估的背景情况（或者原因），目前还不存在对其陈述（Angaben）进行技术的或者法庭—心理学检验的可靠可能性。在此情形，对智能代理—生成的证据的可信性进行严格的查验就变得特别要紧，因为它们们在信任方面享有（也许客观上还没有得到合理化的）优势：社会—心理学研究表明，人们往往赋予机器一种特殊的中立性并由此赋予其日益升高的信用性（Glaubwürdigkeit）⁵⁷。

智能代理由于其学习模式、启发式方法及其数据的标准化，具有典型的局限性和错误源。错误的结论可能是由于不适当的软件—设计、对智能代理的培训不足，以及对数据采集或处理的外部干扰⁵⁸。为了发现这类错误，有必要制定检验智能代理的“诚实性”和信用性的实质标准。但这并不

⁵⁶ 仅见 Barton (Hrsg.), 《言词的但是错误的：质疑证人证言（Redlich aber falsch. Die Fragwürdigkeit des Zeugenbeweises）》，1995；Arntzen, 《证人陈述心理学（Psychologie der Zeugenaussage）》，5. Aufl. 2011；Jansen, 《证人和陈述心理学（Zeuge und Aussagepsychologie）》，2011。

⁵⁷ Mosier/Skitka/Heers/Burdick, International Journal of Aviation Psychology 8 (1997), 47；Cummings, Automation Bias in Intelligent Time Critical Decision Support Systems, 2012 <https://arc.aiaa.org/doi/10.2514/6.2004-6313>。不同的是，在某些情况下，这种有利于机器的偏见（Vorurteil）甚至反映在了法律规则中。例如，在澳大利亚，一旦智能摄像头记录了汽车司机使用手机的情况，就出现一种举证责任的倒置；<www.parliament.nsw.gov.au/bills/Pages/bill-details.aspx?pk=3493>; <theconversation.com/caught-red-handed-automatic-cameras-will-spot-mobile-using-motorists-but-at-what-cost-125638?utm_medium=ampemail&utm_source=email>; <roadsafety.transport.nsw.gov.au/stayingsafe/mobilephones/technology.html>。

⁵⁸ 参见 Cao et al., Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, Association for Computing Machinery, 2267; Takefuji, Cell 2019; Hilgendorf (Fn. 7), 234。

容易，因为它们借助于采集和分析大量的数据，并在高度复杂的数据处理程序基础上，完成其要求很高的任务。如果人们想检验智能代理结论的可靠性，就不应该有智能代理的运作程序与人类观察者所期望的行动方式相一致的看法⁵⁹。在许多情况下，说明智能代理这种（工作）效能的是其结论的正确性（Richtigkeit）和精确性（Präzision）。如果它们由于编程或自学过程而以不切实际的（错误）前提为出发点，或者它们以曲解的方式使用数据，它们当然也可能形成脱离（或不符合）现实的反应⁶⁰。为了克服上述智能代理用于刑事诉讼证据证明而出现的难题，我们提出如下三种解决路径：认证（Zertifizierung），使用鉴定人以及通过智能软件进行查验。

（二）认证

如果人类不能完全理解智能代理的生成和数据处理，那么对它的观察和判断的可靠性的信任或许能够通过校准和持续测定以及认证建立起来⁶¹。这种解决方案的第一个先决条件是制定相关数据范畴的标准：将其存储于DSSAD——智能代理感知和记忆的中心“器官”——的标准以及其随后的可解读标准⁶²。跨学科研究团队可以开发测试目录，借此专业而中立地评价智能代理的技术可靠性（Zuverlässigkeit）并认定其可靠度（Reliabilität）⁶³。根据确证可靠性的流程，应该可以确定，例如，是否可以接受一个驾驶员辅助系统是基本没有故障错误的。确证可靠性或许基于测试数据集的输入，这些数据集必定产生带有特定概率的预期结果（Resultate）。然后，可以确立法律上的规则：只有经过认证的智能代理才可以使用⁶⁴，或者说只有其结论才可以在法庭上予以评价⁶⁵。

当然，即使是智能代理的认证也不能完全解决其结论和评估的有限可解释性问题。例如，如果投入使用驾驶员辅助系统，它会在运行中个别地适应于（或者迎合）具体的人类驾驶员，那么与机器学习和自主反应密不可分的这个问题就会持续存在：即便对于投入市场前经过测试的和认证的辅助系统，也不能排除它会根据其“经验”形成不正确的结论，并在不应该干预的时候进行干预，而人类驾驶员无法确定为什么智能代理得出了错误的结论。此外，总是存在着智能代理——由于未被注意到的安全漏洞而——被外部人员操纵的风险。

（三）鉴定专家的参与

补充使用鉴定专家来协助法官评估智能代理的可靠性似乎颇有希望⁶⁶。专家们可以审查智能代理的设计以及它的机器学习，包括训练数据和使用的的方法，在此不仅可以纳入可能的“偏见（bias, 偏好）”问题，还可以调查其他的，比如智能代理可能受到外部操纵的情况⁶⁷。不过，鉴定者证据对查验智能代理 - 结论的效力取决于几个前提条件：鉴定专家们必须能够接触到有关智能代理收集

⁵⁹ 参见 Roth, Yale Law Journal 126 (2017), 1972, 2006 ff.; Garrett, Cornell Law Review online 99 (2014), 101, 106 f.

⁶⁰ 对此参见上面提到的 Wolf/Husky-Beispiel. 在刑事追诉领域围绕投入人工智能的讨论, s. Hilgendorf (Fn. 7), 232 ff.; Gless StV 2018, 671.

⁶¹ 关于运营商对“高风险”智能代理系统进行认证和持续监测的义务, 参见《人工智能欧洲路径的欧盟条例草案》第13条 (Art. 13 des Entwurfs einer EU Regulation on a European Approach for Artificial Intelligence) (o. Fn. 7). 关于认证对结果可靠性的信任的重要性参见 OLG Köln DAR 2019, 695.

⁶² 关于细节, 参见 VO (EU) 2019/2144 和 ECE/TRANS/WP.29/2020/81.

⁶³ 比如, 参见 <https://www.iais.fraunhofer.de/de/kompetenzplattform-ki-nrw/ki-zertifizierung.html>.

⁶⁴ 这些标准是 VO (EU) 2019/2144 第4条对机动车的驾驶员辅助系统规定的。

⁶⁵ 参见《瑞士刑事诉讼法》第139条第1款：“为了发现真相，刑事司法机关应当使用所有根据科学和经验状况适合的法所允许的证据手段。”

⁶⁶ 参见 Schrey/Haug, NJW 2010, 2917 f. 对此也可参见本文前面三、3、部分。

⁶⁷ Perel/Elkin-Koren, Florida Law Review 69 (2017), 181.

和处理数据的所有相关信息,而且他们必须能够追溯理解智能代理是以何种方式形成其结论的。由于制造商可能想保守他们的商业秘密或者不愿符合标准化的流程,那么这前提条件中的第一个就因制造商的利益而落空(或者不可能具备)。第二个前提条件遇到了上面已经描述过的障碍,即证据的生成——视智能代理的发展阶段不同——可能是基于一个在运行过程中能够自我学习,但目前我们对其学习机制的理解只能有限追溯的系统工程。

(四) 人工反智能是一条出路?

一个新的解决方案可以是力争使用人工智能来查验智能代理的结论。按照对特工活动的反概念的英文名称,我们把这种(未来的)查验软件称为“人工反-智能(ACI: Artificial Counter-Intelligence)”。人工反智能的基本思想是,它检验智能代理的运行可靠性——无论基础技术是什么(基于规则的、人工智能的或者综合的)——而在具体情形下不依赖人类输入指令。

人工反智能可以通过验证错误的概率来给智能代理的性能提供一般说明,从而为法官审查智能代理生成的证据(种类)做好准备。为此,人工反智能或许可以对存储在智能代理的DSSAD中的原始数据和测量数据进行预定义模拟(vordefinierte Simulation)。以这种方式虽然可以对带有其预编程数据处理程序的基于规则的系统,通过相对清晰的程序进行审查,但对于在使用中学习着的系统来说,情况就比较困难了。例如,对于一个车道偏离警示系统,它随着每一公里的行驶和每一种情况的出现而学习,首先必须确定智能代理在有问题的时间点的认知水平,然后必须查验可能的错误或误差。

人工反智能不是科幻,而是设想今天已经存在的、建立在人工智能上的各种方法并不存在。在与法官的沟通中,今天的鉴定专家已经使用概率关系来评估不同的事实假设⁶⁸。人工反智能将为大量的场景形成这种关系,并且以此非常确定地评估智能代理生成的数据的可靠性(Verlässlichkeit)。这就与“值得信赖的人工智能(trustworthy AI)”的研究方法联系起来⁶⁹。

因此,科学面临的任务是生产能够检测各种智能代理行事中可能存在的缺陷和错误的人工反智能。然而,这绝非易事。(开发)带有审查包括其自学程序在内的不同智能代理运作功能的人工反智能的编程,就已经是困难的了。为此,首先有必要对可能的错误源有一个总体的了解。这包括技术应用时的错误、不成功的智能代理培训和有错误的记录。下一步是开展、升级测试和进行模拟-攻击(Schein-Attacken),这些可以显示出智能代理是否有能力防御对其数据处理的攻击⁷⁰。

追问使用人工反智能“审查”智能代理给出的报告,最终导致了多个根本问题:数据处理操作的结果和人类证人的陈词之间是什么关系?⁷¹对智能代理来说,“真实”和“违背真实性”是什么意思(意味着什么)?而如何让“真实”能具体操作呢?⁷²为确保这种证据种类的可靠性(或“可信性”),对人类证人和智能代理能适用同一规则,还是需要为智能代理制定完全独立的诸如遵守ISO(标准化国际组织)或者其他技术标准的规则?

对于人工反智能在刑事程序中的使用,还有一个问题是,如何将调查的结果转化为操作规则,例如智能代理生成的结论的准确性概率必须是多大,才能作为证据(种类)予以评价。最后,这些规范性规则必须转化进入人工反智能的编程。只有当这些步骤成功完成之后,研究人员才能确定人工反智能是否可以用来确定智能代理生成的证据的可靠程度,允许其——类似于其他科学知识——

⁶⁸ Aitken/Nordgaard/Taroni/Biedermann, *frontiers in Genetics* 9 (2018), 1.

⁶⁹ Floridi, *Nature Machine Intelligence* 1 (2019), 261.

⁷⁰ 参见 Mitchell/Mitchell/Mitchell *Law, Probability and Risk* 19 (2020), 43.

⁷¹ 参见 Gless, *Georgetown Journal of International Law* 51 (2020), 195.

⁷² 对此,参见 Anglano/Canonico/Guazzone, *Technical Report* 2017, 1 ff.; Roth, *Yale Law Journal* 126 (2017), 1972, 2005 ff.; Chessman, *California Law Review* 105 (2017), 179; Hilgendorf (Fn. 7), 245 ff.

在刑事程序中使用⁷³。在这个领域，法律人和智能代理技术专家们之间有必要进行紧密的合作。法律人的任务是研究出在刑事程序中使用证据种类的规范性前提。而智能代理专家们必须将这些法上的原则与智能代理的启发式决策和自我学习模式联系起来⁷⁴。在这些考虑的基础上，才可能制定出智能代理和人工反智能对于刑事程序的充分“可靠性”的一般标准。

对智能代理 - 证据和人工反智能（ACI）审查之间的程序性联系来说，可以想到多个可能。第一种可能，只有当智能代理生成的结论预先已经受到人工反智能的审查，人们才能允许它作为证据。只有当这种审查显示出一定的——法律上预先规定的——可靠性水平时，才会允许它们作为证据种类进入诉讼程序。接下来将由法官决定如何评估该证据种类以及在多大程度上把它作为判决之根据或基础。第二种可能或许更有利于智能代理，原则上允许智能代理生成的证据进入诉讼程序，但是要赋予法官和程序参与者准许其进行人工反智能审查的权力。接下来该审查结论同样作为证据（种类）可以进入诉讼，法官必须对证据状况进行（自由）心证的总体考量。第二种规则安排的好处是可以加快程序，减少费用支出，如果任何一方对智能代理生成的证据（种类）没有提出重大的反对意见，那这就特别富有意义。

回答所有这些问题是一项艰巨的任务。但这个目标是值得努力的。因为使用人工反智能可以克服上面概述的智能代理 - 证据在证据法上的两个难题：人工反智能可以确保智能代理决策过程的理性足以形成法官确信；人工反智能审查至少可以提供相当于被告人对不利证据种类的“对质”功能。这样一来，刑事程序法的传统准确性保障就与在程序中——而且不仅仅是刑事程序中——使用智能代理生成的证据种类寻找真相的潜力结合了起来⁷⁵。

六、总结

智能代理越来越多地伴随并方便着我们的生活。与此同时，它们持续地观察着我们和我们周围的环境，并收集着大量的、也可用于刑事程序查明案件事实的数据。目前，还不能把智能代理对人类行为的评估引入严格的证据程序，因为它还没有被列入公认的证据种类目录中。因此，如果要把智能代理观察和行动的结论作为证据来使用，就必须专门扩展证据种类的目录。

当然，在此——除了必要的修法之外——还有许多需要解决的问题，同时在一定程度上要修改我们针对人类证人的传统证明程序。在刑事程序的框架内，必须创造可能性来查验感知的可靠性。这里的问题是，一方面，智能代理不能展现出他们的人格信用性，另一方面，我们事后并不总是能够完全弄清楚它们形成反应的路径。这就在证据证明的直接性、事实调查过程的理性以及被告人对质、查验对其不利的证据的权利（《欧洲人权公约》第6条第3款d项）方面出现了问题。如果人们想获得这些问题的解决，就必须看到，智能代理和人类在功能运作上是不同的，信用性的检验只能在各自的物种（Gattung）内有效进行。因此，为了确保借助智能代理进行证据证明的理性，建议开发能够查验智能代理的运作及决策路径的软件（人工反智能：ACI），并将这种查验的结论提供给法官和程序参与者。

（收稿：2024-05-05，修回：2024-07-01）

（责任编辑：）

⁷³ 法庭科学通常可以给出不同的可想象的事件序列之间的概率关系；参见 *Evetts et al. Journal of Forensic Sciences* 47 (2002), 520; *Aitken/Nordgaard/Taroni/Biedermann, frontiers in Genetics* 9 (2018), 1.

⁷⁴ 比如，参见 *Završnik, ERA Forum* 20 (2020), 567.

⁷⁵ 同时，制定超越国内刑事程序法规则的标准，也可以促进欧盟内部程序基本规则的统一，从而改善欧洲的司法互助。对此，参见 *Kirmes, Risk Fraud & Compliance* 2013, 66 ff.; *Mitsilegas*, in: *Brown/Turner/Weisser (Hrsg.), The Oxford Handbook of Criminal Process*, 2019, 115; *Tosza, New Journal of European Criminal Law* 11 (2020), 161.