Zeitschrift für die gesamte Versicherungswissenschaft, Published Online First
https://doi.org/10.3790/zverswiss.2025.1469701

Safeguarding Medical Decision Quality in the Age of AI – A Practice Report from Medical Risk and Claims Assessment in a Life Insurance Company

Jens Rickmann and *Verena Tröster*

Abstract

The increasing use of artificial intelligence (AI) in the risk and claims assessment of biometric life insurance products necessitates new qualification profiles. While automated decision systems promise efficiency gains, expert human judgement of complex medical conditions remains essential. This article outlines the ethical, regulatory, and professional requirements for a "human-in-the-loop" approach in AI-assisted decision-making. It then presents a company-internal qualification program designed to systematically strengthen medical competence among non-physician underwriters and claims assessors. The modular, hybrid-format training curriculum comprised 105 teaching hours, including 23 practice-oriented case conferences. Standardized evaluation revealed very high participant satisfaction and a substantial gain in medical judgement capabilities. Participants highlighted the program's didactic structure, practical relevance, and collaborative atmosphere. The results demonstrate that medical qualification can be effectively and systematically fostered outside formal licensing frameworks. The training format has since been integrated into a continuous qualification program that also serves as a foundation and ongoing support for the increasing use of AI in risk and claims assessment.

Corresponding Author: Dr. med. Jens Rickmann MPH,
Chief Medical Officer and Company Physician,
Alte Leipziger Lebensversicherung a.G.,
Alte Leipziger-Platz 1,
61440 Oberursel,
Germany,
jens.rickmann@alte-leipziger.de.

Verena Tröster M.Sc. Psychology,
Alte Leipziger Lebensversicherung a.G.,
Alte Leipziger-Platz 1,
61440 Oberursel,
verena.troester@alte-leipziger.de

Duncker & Humblot · Berlin

Zusammenfassung

Der zunehmende Einsatz von Künstlicher Intelligenz (KI) in der Risiko- und Leistungsprüfung biometrischer Lebensversicherungsprodukte erfordert neue Qualifikationsprofile: Während automatisierte Entscheidungssysteme Effizienzgewinne versprechen, bleibt die fachlich fundierte menschliche Bewertung komplexer medizinischer Sachverhalte unverzichtbar. Der vorliegende Beitrag skizziert die ethischen, regulatorischen und fachlichen Anforderungen an den „Human-in-the-Loop"-Ansatz in KI-gestützten Entscheidungsprozessen. Anschließend wird ein unternehmensintern entwickeltes Curriculum zur versicherungsmedizinischen Qualifikation nicht-ärztlicher Prüfpersonen vorgestellt, das sowohl medizinisches Grundlagenwissen als auch Anwendungskompetenz systematisch fördert. Das modular aufgebaute, hybrid durchgeführte Fortbildungsprogramm umfasst 105 Unterrichtsstunden, darunter 23 praxisnahe Fallkonferenzen. Die standardisierte Evaluation zeigte eine sehr hohe Zufriedenheit der Teilnehmenden sowie einen substanziellen Lerngewinn im Bereich medizinischer Urteilskompetenz. Besonders hervorgehoben wurden die didaktische Struktur, der Praxisbezug sowie der kollegiale Austausch. Die Ergebnisse belegen: Medizinische Qualifikation lässt sich auch außerhalb formaler Berufszulassungen wirksam und strukturiert fördern. Das Schulungsformat wurde aufgrund dieser Erkenntnisse mittlerweile in ein kontinuierliches Qualifikationsprogramm überführt, auch als Grundlage und Begleitung der zunehmenden Nutzung von KI in der Risiko- und Leistungsprüfung.

1. Introduction

The use of artificial intelligence (AI) is transforming operational practices across a wide range of industries – including the healthcare and insurance sectors. In the risk assessment and claims evaluation of biometric insurance products, such as disability, life, or other invalidity insurance, AI-supported systems are no longer a promise of the future; they are increasingly becoming part of operational reality. Examples include automated pre-selection processes, natural language processing for extracting medical data, and risk-based decision support systems in underwriting and claims handling.

The anticipated benefits are substantial: faster decisions, reduced costs, and greater consistency. However, these technological capabilities also entail growing responsibility. Medical assessment processes are not purely formal procedures; they require contextual understanding, ethical consideration, and insurance medical judgement. The World Health Organization (WHO 2024), in its global guidelines on the use of AI-supported systems, emphasizes that humans must retain control over medical decision-making and that human oversight must remain an integral component of any AI application.

2. Problem Statement

The increasing automation of decision-making processes poses a paradoxical risk: the more AI systems are integrated into these processes, the greater the temptation to delegate responsibility to them. This concern was already highlighted by the Zentrale Ethikkommission bei der Bundesärztekammer (2021), the Central Ethics Committee of the German Medical Association, which – like the WHO – emphasizes the principle of final professional responsibility. Physicians must not rely blindly on AI systems but are required to assess the plausibility of AI-generated results based on their training and professional expertise.

This position was further elaborated by the Bundesärztekammer (German Medical Association) (2025), which states that automated decision proposals may be based on non-transparent processes. Therefore, medical expertise is essential – not only for plausibility assessment but also for communicating risk to affected individuals.

The European legislator has also acknowledged the need for regulation. In Annex III, Section 5 of the EU AI Act (Verordnung (EU) 2024/1689), AI systems used for premium setting, risk assessment, or claims evaluation in the context of life and health insurance are explicitly classified as "high-risk AI." These systems are subject to stricter requirements, particularly with respect to the qualifications of their users.

Article 14 of the Regulation defines detailed requirements for human oversight in the use of high-risk AI systems, with far-reaching implications for their practical application – especially in the healthcare and insurance domains. The central implication is that such systems may not be deployed in a fully autonomous manner. Rather, they must be designed and configured in such a way that qualified individuals are always able to supervise their operation, detect anomalies, critically review outputs, and intervene or override decisions if necessary. Of particular importance is the requirement to actively counteract automation bias – that is, the tendency to uncritically trust AI-generated outputs without subjecting them to expert scrutiny.

This gives rise to two distinct sets of competency requirements:

- AI-specific: Knowledge of the system's functioning, limitations, explainability, and potential risks;
- Domain-specific: the ability to interpret the system's output in its applied context – e. g., medical issues such as diagnoses, ICD codes, or treatment trajectories. While the latter is not explicitly required by law, it follows implicitly from the scope of responsibility and the operator's legal accountability.

In life insurance practice, however, it is typically not physicians who are responsible for risk assessment or claims processing. Rather, these tasks are per-

formed by insurance professionals or staff with backgrounds in nursing, therapy, or medical-technical fields. Accordingly, the ethical and professional standards laid out in the aforementioned policy papers and empirical studies – regarding medical judgement, accountability, and plausibility checking – apply by analogy to this group of non-physician professionals. This is especially relevant given that these individuals routinely perform extensive medical assessments, make preliminary decisions, and engage in professional dialogue with external medical experts and assessors.

The need to systematically train non-physician claims and risk assessors in the professional handling of AI-assisted tools for health-related decisions arises not only from a concern for technical proficiency but also from regulatory obligations.

In their market study, Salgaonkar, Allen and Zimmerman (2022) show that, despite increasing automation, company physicians in life insurance continue to play a key role in complex cases. At the same time, the study reports that only around half of the companies surveyed invest substantially in the continuing education of medical professionals. This suggests a growing discrepancy between increasing case complexity and stagnating professional qualification – a discrepancy that may constitute a material operational risk.

This challenge is also reflected in empirical studies on the cognitive impact of generative AI. In a large-scale study by Lee et al. (2025) involving 319 knowledge workers, researchers found that greater trust in GenAI correlated with a reduced willingness to engage in critical thinking, whereas those with greater confidence in their own professional competence were more likely to question AI outputs.

The reinsurer Gen Re (Schilling/Eppert 2024) similarly emphasizes that the final evaluation of an application must remain the responsibility of a qualified medical underwriter – even in the context of increasing automation. A full delegation of decision-making is considered neither legally nor professionally acceptable.

The reinsurer Munich Re (Sarkin 2025) shares this view. In a presentation on the concept of AI-Augmented Underwriting at the 2025 Data Analytics Virtual Forum (DAVF) hosted by the International Actuarial Association, the company stressed that highly complex cases require deep domain expertise and that qualified personnel must be actively involved in managing the interface between automation and risk analysis.

Finally, an ethics-oriented analysis by Mullins, Holland and Cunneen (2021) on the use of AI in the European insurance industry concludes that explainability and traceability of algorithmic decisions can only be ensured through the involvement of sufficiently qualified professionals. This, they argue, requires a

combination of technical understanding and insurance medical contextual competence.

3. From Principle to Practice – A Qualification Program from the Field

The previously outlined ethical, regulatory, and professional requirements for the use of AI-supported systems in medical risk assessment and claims evaluation make one thing clear: the role of medically trained examiners is not being diminished – on the contrary, it is becoming more complex, more responsible, and more integral to decision-making. The ability to contextualize, critically assess, and interpret machine-generated information is a key competence in the age of artificial intelligence. This applies equally to underwriting and claims handling – especially for products that require a high degree of medical evaluation, such as disability insurance.

This finding is not merely theoretical but has had concrete implications for the company in which the authors of this article are employed – a life insurance provider specializing in biometric products. In 2022, as part of the Medical Department, which is also responsible for medical quality assurance, we initiated a structured qualification program for professionals involved in medical risk assessment and claims evaluation. The program aims to systematically strengthen the medical judgement of our staff and to establish a critical foundation for the emerging interdisciplinary integration of AI-supported tools in the years ahead.

At the heart of the initiative lies the transfer of competencies necessary for the sound evaluation of complex medical cases – both on the risk side during application assessment and on the claims side for challenging disability insurance cases. The objective is to enable our examiners not only to interpret medical information with confidence but also to critically assess, validate, and, if necessary, correct AI-generated suggestions – following the "human-in-the-loop" principle.

The qualification program was designed as an insurance medical curriculum that combines a needs-based, comprehensive medical knowledge transfer with a clear focus on insurance-relevant medical questions. It specifically addresses the requirements of medical risk and claims assessment in the context of biometric life insurance products.

4. Thematic Structure of the Curriculum

The training program was structured into the chapters listed in the Table.

Table

Overview of Curriculum Structure

No.	Chapter	Instructional Hours	Presentation Slides
0	Introduction	1	8
1	Fundamentals of Medical Risk and Claims Assessment	1	10
2	Medical Diagnostics and Reporting	6	49
3	Body Structure and Heredity	1	16
4	Blood Pressure	1	17
5	Heart	7	107
6	Vascular System	2	45
7	Blood	2	44
8	Hormones	2	32
9	Diabetes Mellitus and Metabolism	3	34
10	Digestive Organs	7	117
11	Urogenital Organs	5	91
12	Neurology	6	111
13	Mental Health Disorders	3	42
14	Tumor Diseases	6	127
15	Lungs	4	78
16	Infectious Diseases	4	95
17	Orthopedics I – Upper Extremities	4	64
18	Orthopedics II – Lower Extremities	4	90
19	Orthopedics III – Spinal Disorders	3	49
20	Eyes and Ear-Nose-Throat (ENT)	5	74
21	Dermatological Conditions	2	35
22	Immune System	3	80
	Total (excluding case conferences)	**82**	**1,415**

In addition to these 22 thematic chapters, structured case conferences were held following chosen modules – separately for risk assessment and claims evaluation. These practice-oriented, interactive sessions enabled a differentiated application and reflection of the previously conveyed content within the relevant professional context. The total time allocated to the case conferences amounted to an additional 23 instructional hours, bringing the full scope of the qualification program to 105 teaching units.

5. Didactic Format

The training units were generally conducted as two one-hour meetings per week in a hybrid conference format – meaning that participants could attend either in person or remotely via video conferencing. This flexible structure enabled consistent participation even within the day-to-day demands of operational assessment work.

The program was designed as a continuous, modular learning format, systematically combining content delivery during chapter-based instructional sessions with applied reflection through the transfer of acquired medical knowledge to concrete case evaluations during subsequent case conferences. This approach established a framework that equally promoted both foundational medical knowledge and its application in insurance-specific assessment contexts.

Medical instruction was delivered through structured presentations, developed and personally presented by the two instructors (the authors of this article). The content was tailored specifically to the requirements and perspectives of medical risk assessment and claims evaluation for biometric life insurance products. Each session included opportunities for participant questions and deliberately referenced real-life medical assessment situations in underwriting and claims processing.

Following chosen instructional units, case conferences were conducted – separately for risk and claims assessment. The objective of these sessions was to anchor the previously covered medical content within its relevant application context, reinforce assessment confidence, and make differences in individual evaluative approaches transparent.

To promote knowledge transfer and practical assessment competence, the instructors intentionally adopted a facilitative and observational role during the case conferences. The responsibility for content – particularly the selection, preparation, and presentation of cases, as well as the moderation of discussions – rested entirely with the participants. This created a learning environment conducive to collegial evaluation practice, in which shared standards could be developed and individual uncertainties addressed constructively.

6. Evaluation

To ensure standardized feedback and quality assurance, a comprehensive evaluation questionnaire was administered at the conclusion of the training program. The questionnaire was specifically developed for this initiative and comprised nine thematic categories with approximately 30 individual items, along with open comment sections. Most items were assessed using five-point Likert scales ranging from "*very good*" to "*very poor*" or "*always*" to "*never*".

The thematic categories included:

1. Scope of participation and interest in the topics
2. Clarity and structure of seminar planning
3. Learning environment and support provided by the instructors
4. Motivation for active participation
5. Integration of academic and scientific perspectives
6. Teaching methodology and use of media
7. Practical relevance and application orientation
8. Observed participant engagement and preparation
9. Individual learning gains and impact on professional practice

The questionnaire was designed to yield both quantitative assessments of the overall quality of the program and qualitative feedback on individual modules, teaching methods, and areas for improvement.

7. Evaluation Results

The voluntary and anonymized evaluation of the insurance medical training program achieved a response rate of 100 % (n = 20) – a strong indicator of the high level of participant identification with the program. Quantitative scores across nearly all evaluation categories were consistently positive to very positive, reflecting a high level of satisfaction with the program's structure, content, and didactic approach. The average ratings across all evaluation categories are presented in the Figure.

In terms of content, the feedback emphasized a significant increase in learning, particularly in the area of medically functional knowledge. Several participants reported experiencing distinct "a-ha" moments especially when gaining a deeper understanding of complex medical conditions such as cancer and immunological disorders for the first time.

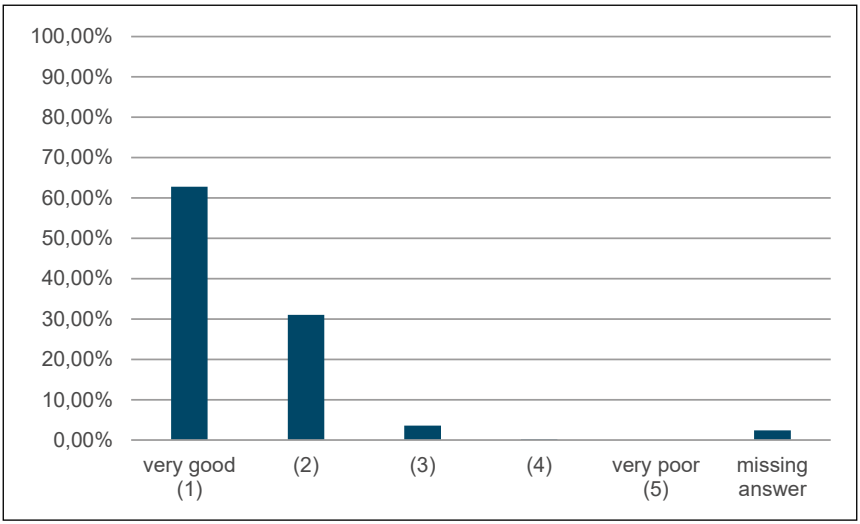

Figure: Average rating of evaluation questionnaire items for the insurance medical training program by n = 20 participants

This was accompanied by a noticeable improvement in evaluation competence. The training enabled participants to interpret medical findings more reliably, assess functional limitations more precisely, and better understand the relationship between medical diagnostics and occupational performance. This, in turn, was associated with a heightened self-perception of argumentation skills and independent decision-making.

The didactic structure of the program also received special praise. Complex subject matter was conveyed in an accessible and comprehensible manner, supported by visual tools such as anatomical software and structured slide decks. At the same time, the feedback clearly indicated that such content requires a carefully designed instructional framework in order to be effective in the long term.

Participants assigned particular value to the interactive case conferences. These sessions not only deepened subject-matter knowledge but also facilitated the practical transfer of learning into the insurance medical context – especially with regard to the evaluation of complex claims cases.

8. Lessons Learned and Outlook

The evaluation results clearly demonstrate that medical expertise remains an indispensable foundation for quality-assured decisions in risk and claims assess-

ment – even in an era of algorithmic decision support. The observed enhancement of medical judgement skills not only benefits manual case handling but also becomes a prerequisite for the informed and responsible use of AI systems.

The ability to critically evaluate AI-generated suggestions, assess their plausibility, and responsibly incorporate them into decision-making processes requires deep medical understanding – particularly in domains characterized by complexity, ambiguity, and sensitivity to context.

Moreover, the evaluation underscores the strategic importance of methodically structured training as a response to technological change. Educational formats such as case conferences, which interweave medical knowledge with insurance-specific contexts, serve as exemplary models of how human expertise and data-driven decision logic can be meaningfully integrated.

The key insight: The introduction of AI does not replace medical competence – it amplifies its relevance. Only qualified professionals with medical backgrounds can ensure the quality, traceability, and ethical integrity of future decision-making processes.

The evaluation findings and practical experience also show that medical qualification can be systematically and effectively fostered outside of formal professional licensure, provided that structure, relevance, and practical applicability are consistently addressed. The training program has shown that a well-designed curriculum can enable even non-medically licensed participants to competently understand and assess complex medical information within insurance-related contexts.

The following factors were identified as particularly decisive for the success of the format:

- The depth and professional medical orientation of the seminar content
- The ongoing moderation and instructional support by medical and psychological professionals
- The clear methodological separation between knowledge acquisition and practical application
- The principle of collegial responsibility during case conferences
- And the integration into everyday operations through fixed time slots, hybrid formats, and strong alignment with professional routines.

Despite operational challenges, the training series was fully and consistently implemented. Participant feedback confirmed not only its direct utility for improving the quality of risk and claims assessment, but also its indirect effects on motivation, collegial exchange, and the development of shared medical standards within the examiner team.

In light of the positive evaluation and strategic relevance, the training concept has been transitioned into a continuous learning loop. This means that the program is conducted on an ongoing basis and is regularly updated and further developed in both content and structure. The participant base is continuously expanded – to include new employees, selected professionals with diverse academic or professional backgrounds (e. g., from law or related assessment domains), as well as experienced risk and claims assessors brought in for targeted enrichment and peer knowledge exchange.

Looking ahead, the transfer of the training concept to other product lines – such as private health insurance – appears both logical and feasible. Many of the methodological elements (e. g., problem-oriented knowledge transfer, case-based learning, hybrid instructional formats) can be adapted with minimal effort and may contribute to establishing a structured and practice-relevant medical qualification profile in other areas of insurance.

Accordingly, our qualification program serves not only as a tool for methodically structured internal training, but also as a strategic lever in managing the digital transformation of life insurance.

References

Bundesärztekammer (2025): Stellungnahme: Künstliche Intelligenz in der Medizin. Deutsches Ärzteblatt, 122(4), 1 – 14.

EU (2024): Verordnung (EU) 2024/1689 des europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz). ABl. L, 2024/1689, 1 – 144.

Gen Re (*Schilling*, A./*Eppert*, T.) (2024): KI in der Risikoprüfung [Konferenzbeitrag]. Kongress Versicherungsmedizin, Köln, Germany. 20.11.2024.

Lee, H.-P./*Sarkar*, A./*Tankelevitch*, L./*Drosos*, I./*Rintel*, S./*Banks*, R./*Wilson*, N. (2025): The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers [Konferenzbeitrag]. CHI Conference on Human Factors in Computing Systems (CHI '25), Yokohama, Japan. https://doi.org/10.1145/3706598.3713778.

Mullins, P./*Holland*, C. P./*Cunneen*, M. (2021): Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the European insurance market. Patterns, 2(10), Article 615826.

Munich Re (*Sarkin*, L.) (2025): AI-Augmented Underwriting: Balancing Benefits and Risks [Konferenzbeitrag]. Data Analytics Virtual Forum (DAVF) 2025 der International Actuarial Association, Ottawa, Canada. 24.04.2025.

Salgaonkar, S. S./*Allen*, L./*Zimmerman*, D. (2022): Checking the Pulse: A Global Survey on the Evolving Roles of the Insurance Medical Director. Journal of insurance medicine, 49, 189 – 199.

WHO – World Health Organization (2024): Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. Genf.

Zentrale Ethikkommission bei der Bundesärztekammer (2021): Entscheidungsunterstützung ärztlicher Tätigkeit durch künstliche Intelligenz. Deutsches Ärzteblatt, 118(33/34), A1 – 13.