ONE80 INTERMEDIARIES

Brokers' Guide to Artificial Intelligence

AI Warranty Coverage

Innovating at the pace of change

Table of Contents

To say that AI is the topic of the year is a huge understatement, for good reason. A recent study from McKinsey estimates that Generative AI (GenAI) could add the equivalent of $2.6 trillion to $4.4 trillion of annual value to our economy. At the same time, the public debate whether AI is ultimately net positive or net negative will rage on for years and will vary based on perspective. The rate at which AI has come into the public consciousness is astounding. It is nearly impossible for any single company to completely grasp the downstream effect of implementing or not implementing AI tools into their business. In fact, for most businesses, their vendors are utilizing AI (i.e Customer Service ChatBots) even if the business itself isn't actively engaging AI.

The speed of change and the massive financial implications require that there is a reliable risk management mechanism which can evaluate the risk of any given AI product and transfer the risk, for those companies that so desire.

What is AI?

In its most basic form, Artificial intelligence, or AI, is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities. On its own or combined with other technologies (e.g., sensors, geolocation, robotics) AI can perform tasks that would otherwise require human intelligence or intervention.

Use Cases *(THIS SECTION WAS CREATED WITH THE HELP OF PERPLEXITY)*

At a high level, Digital assistants, GPS guidance, autonomous vehicles, and generative AI tools (like Open AI's Chat GPT) are a few examples of commonly used AI models.

Some more specific examples of AI at work include:

Healthcare

- AI models can be used to read x-rays and identify specific types of disease or injury.
- AI is used to support patient triage and risk classification.
- AI is used to transcribe procedure payment codes and automate medical billing.
- Clinicians use AI-powered tools to digitize, summarize and analyze medical notes.
- Telehealth businesses can use AI to remotely monitor patients.
- During a drug or device trial AI can be used to speed up the approval process by running thousands of simulations

Human Resources

- Software uses AI to scan resumes and filter the best applicants.
- Talent or skills management platforms are used to help employers retain, upskill or redeploy employees.
- Performance reviews can utilize AI to assist in the completion of the assessment and recommend aspects of the employees' career path including continuing education and readiness for promotion.

Automation

- Self-driving cars use AI to identify obstacles and follow traffic rules.
- AI can be used to predict when equipment and vehicles will need maintenance.
- Self-driving vehicles rely on large quantities of both human-labeled and AI-labeled data to train their AI systems to drive in routine and extreme conditions.

Financial Services

- Credit scoring uses complex neural networks to inform their lending decisions.
- Banks use detect fraud using AI algorithms by analyzing vast amounts of data in real time.

Legal Services

- In-house counsel and law firms rely on AI-powered tools to assist with contract drafting, and to automate red-lining with improved accuracy and efficiency
- New generative AI tools exist to support drafting of legal memos and claims

Insurance

- Insurance carriers rely on AI-drive tools to create efficiencies in terms of speed and accuracy in underwriting, risk assessment, claims management, fraud detection and pricing.

Enterprises across telecommunications, customer service, banking and financial services, insurance and retail are relying on Generative AI software to enhance the customer experience, through support agents, personal assistants, copilots and more.

All of these use cases represent unmistakable benefits to our economy with the potential to make life better for billions of people. At the same time, we are also starting to see what can go wrong with AI. In fact, just last month the 2024 edition of the Stanford HAI Index, a leading authority on trends and impacts in the AI industry, as well as research and policy, highlighted that AI incidents were on the rise – having increased more than 25 times incident rates collected in 2012.

Prominent AI failures:

Microsoft's Chatbot Tay

Microsoft's AI chatbot Tay was designed to engage in conversations on Twitter. However, it quickly began to post racist and offensive content after learning from interactions with users who fed it inappropriate language. This incident underscored the risks of deploying AI systems without robust safeguards against learning harmful behaviors from user interactions.

Amazon's AI Recruiting Tool

Amazon developed an AI-powered recruiting tool intended to streamline the hiring process by analyzing resumes. Unfortunately, the tool exhibited gender bias, favoring male candidates over female ones. This bias stemmed from the training data, which predominantly featured resumes from men, reflecting existing gender imbalances in the tech industry. A mazon had to shut down the tool after its biased nature was revealed.

Google Photos Mislabeling

Google Photos' algorithm once incorrectly labeled photos of people with darker skin tones as "gorillas." This incident highlighted the significant issue of racial bias in AI systems and the importance of diverse and representative training data to avoid such egregious errors.

Zillow's Home-Buying Algorithm

Zillow's AI-driven home-buying algorithm, known as "Zestimate," led to significant financial losses and the layoff of 2,000 employees. The algorithm's predictions on home prices were inaccurate, causing the company to overpay for homes and subsequently sell them at a loss. This failure demonstrated the potential economic risks of relying heavily on AI for critical business decisions without adequate oversight and validation.

Deepfake Scams

Deepfake technology has been used to create highly realistic but fake audio and video content. In one notable case, scammers used an audio deepfake to impersonate the CEO of an energy company, convincing a subordinate to transfer hundreds of thousands of dollars. This incident illustrates the potential for AI to be used in sophisticated fraud schemes, posing significant security risks.

AI in Social Media

AI algorithms used by social media platforms to maximize user engagement have been criticized for contributing to the spread of misinformation and creating echo chambers. These algorithms often prioritize sensational or emotionally charged content, which can distort public discourse and exacerbate social divisions.

IBM Watson for Oncology

IBM's Watson for Oncology was intended to assist doctors in diagnosing and treating cancer. However, it often provided incorrect or unsafe treatment recommendations, leading to its failure in clinical settings. This case highlights the critical need for rigorous testing and validation of AI systems in healthcare to ensure patient safety. These examples underscore the importance of robust AI safety measures, ethical considerations, and continuous monitoring to mitigate the risks associated with AI technologies.

NTD

Could include prominent AI-bias related class actions against companies such as State Farm, United Health, Cigna in insurance, or against HR Tech companies such as WorkDay. There is also the case of the Dutch Government, which was toppled over its tax authority's use of an inaccurate and biased fraud detection system, which erroneously flagged tens of thousands of families, often lower income or belonging to ethnic minorities, of suspected flawed.

Do Other Policies Cover AI?

Tricky question because the adoption of AI is moving quickly. It is safe to say that the current product, AI Warranty, is not likely covered by other products in the same way. the AI Warranty reimburses the client via a parametric payout if the model is not performing to the agreed level. The parameters are better spelled out in the coverage section below.

As for "liability", the market is very much in flux. Underwriters will generally say that they are underwriting to the exposure, but that just means that if the underwriter thinks there is significant exposure they will likely exclude "AI Losses".

As we have seen when other new policies come into the markets (Employment Practices Liability & Cyber most notably) the legacy products do not initially exclude coverage, but the underwriters start to get very stringent in defining the scope of coverage and eventually exclude it all together. There is no sign that this time will be any different.

Cyber – Most Cyber policies include coverage for "Media Liability" which generally covers infringement of copyright or trademark, invasion of privacy, libel, slander, plagiarism, or negligence by the Insured for content on their website. The Cyber policy doesn't address errors in the data used to train the model unless there was a "bad actor" involved. AI can also drive an adjacent but different set of exposures that need to be addressed in the Cyber policy or elsewhere. For example, traditional cyber security policies don't fully address all of the threats posed by AI, such as data poisoning attacks, prompt injection attacks or the misuse of AI in generating information.

Tech E&O – In the current market, Tech E&O can address many AI exposures, but there are definite exclusions that need to be addressed, depending on the AI powered product. For example, a Tech E&O policy will likely exclude losses arising from Bodily Injury or Property Damage. For companies whose business involves human interaction with machines (i.e. autonomous vehicles) this can severely limit coverage. Similarly, non-compliance with developing AI regulations can lead to claims. The evolving regulation related to AI just makes the situation more onerous.

Directors and Officers Liability – At the Corporate level the D&O policy can respond to AI related claims, for example a Securities Class Action can allege that a company made false or misleading statements about their AI capabilities. However, like the other policies, there are a number of tricky issues that need to be addressed. A broadly worded cyber exclusion in the D&O policy can apply to the use of various models, including AI. The Professional Services Exclusion with broad lead in language ("based upon, arising out of, or attributable to") can create a nearly universal exclusion for AI, in certain situations.

General Liability – Covers claims for bodily injury, property damages and personal/advertising injury arising from the Insured's products or operations. This could extend to a defective AI model, but a thorough coverage review is needed. General Liability policies are not specifically intended to insure AI risks, so it is important to review the specific terms for grants of coverage and exclusionary language.

What Does the AI Warranty Model Cover?

NOTE: A assessment of the model complexity and risk must be performed by Armilla before any warranties can be issued.
NOTE: The information is a summary of the terms and conditions.

1. Maximum limit per warranty contract - $1,000,000
2. Current Limit per model (e.g. if the same model is used in products sold to several clients) - $3,000,000
3. Maximum limit per client - $3,000,000
4. Warranty - Based on our confidence in the effectiveness of the AI Verification Services, as specified in the AI Verification Report that we provide to the Vendor, we hereby provide this Software Performance Warranty to you that the Vendor's services will perform in accordance with the performance metrics set forth in the assessment.
5. Exclusions - The Software Performance Warranty also does not apply to and specifically excludes any of the following:
    - Use of the Vendor's model with missing, inaccurate, incomplete, or misleading data relative to the training and test data evaluated by Armilla AI as specified in the AI Verification Report.
    - Use of the Vendor's model that is inconsistent with the Vendor's Terms of Service, inconsistent with the AI Verification Report, or intended use as specified in Appendix A.
    - Use of the Vendor's model that falls within the specified limitations in the Vendor's Terms of Service, the AI Verification Report, or limitations as set forth in Appendix A.
    - The user's inability to access the Vendor's model due to the unavailability of the system(s) that host the model and provide access to the model.
    - Failures of the model resulting from a cyber-attack, adversarial attack, or an incident that is covered by your cyber-insurance policy.

- The Vendor's modification to the Vendor's model that changes the performance relative to the AI Verification Report or evaluation of the Vendor's performance metrics as set forth in Appendix A.
- Failures of the model arising from the occurrence of a force majeure event, including without limitation earth movement; fire; disease; war; nuclear reaction, radiation or contamination; or terrorist activity; or any activity involving the use or threatened use of any nuclear, biological, chemical, or radioactive agent, material, device or weapon excluding nation state cyber-attacks. Furthermore:
- The Software Performance Warranty does not include faults caused by pre-processing or post processing steps outside of the scope of your agreement with the Vendor or the Vendor's Terms of Service, as applicable, or the AI Verification Report set forth in Appendix A.
- Armilla AI is only warranting the effectiveness and accuracy of the AI Verification Services provided to the Vendor (resulting AI Verification report and evaluation for the specified use case and metrics specified in Appendix A) and of which you are the beneficiary of through the use of the Vendor's model.
- Armilla AI is not responsible for anything beyond the reimbursement provided pursuant to these Software Performance Warranty Terms, and only based on the payout schedule in the Appendix A.
- Armilla AI is not responsible for misuse or misrepresentation of the Software or any other factor beyond its control.
- Any faults arising from government and/or regulatory actions are excluded.

6. Reimbursement - You may file a reimbursement request directly with Armilla AI at www.armilla.ai/warranty/claim after first notifying the Vendor as specified in Appendix A, and allowing the Vendor a remediation period as specified in Appendix A. Such reimbursement request shall include the following information: date period of the claim, date of remediation period start and end, data required to reproduce the results, access to the version of the model used in production and the expected results. Armilla AI will validate the reimbursement request and directly pay you. Armilla AI will complete its processing of any reimbursement request within a reasonable period following the date we have contacted.

AI Glossary of Terms *(source CNET)*

Artificial General Intelligence, or AGI: A concept that suggests a more advanced version of AI than we know today, one that can perform tasks much better than humans while also teaching and advancing its own capabilities.

AI ethics: Principles aimed at preventing AI from harming humans, achieved through means like determining how AI systems should collect data or deal with bias.

AI safety: An interdisciplinary field that's concerned with the long-term impacts of AI and how it could progress suddenly to a super intelligence that could be hostile to humans.

AI safety: An interdisciplinary field that's concerned with the long-term impacts of AI and how it could progress suddenly to a super intelligence that could be hostile to humans.

Algorithm: A series of instructions that allows a computer program to learn and analyze data in a particular way, such as recognizing patterns, to then learn from it and accomplish tasks on its own.

Alignment: Tweaking an AI to better produce the desired outcome. This can refer to anything from moderating content to maintaining positive interactions toward humans.

Anthropomorphism: When humans tend to give nonhuman objects humanlike characteristics. In AI, this can include believing a chatbot is more humanlike and aware than it actually is, like believing it's happy, sad or even sentient altogether.

Artificial intelligence, or AI: The use of technology to simulate human intelligence, either in computer programs or robotics. A field in computer science that aims to build systems that can perform human tasks.

Bias: In regards to large language models, errors resulting from the training data. This can result in falsely attributing certain characteristics to certain races or groups based on stereotypes.

Chatbot: A program that communicates with humans through text that simulates human language.

ChatGPT: An AI chatbot developed by OpenAI that uses large language model technology.

Cognitive computing: Another term for artificial intelligence.

Data augmentation: Remixing existing data or adding a more diverse set of data to train an AI.

Deep learning: A method of AI, and a subfield of machine learning, that uses multiple parameters to recognize complex patterns in pictures, sound and text. The process is inspired by the human brain and uses artificial neural networks to create patterns.

Diffusion: A method of machine learning that takes an existing piece of data, like a photo, and adds random noise. Diffusion models train their networks to re-engineer or recover that photo.

Emergent behavior: When an AI model exhibits unintended abilities.

End-to-end learning, or E2E: A deep learning process in which a model is instructed to perform a task from start to finish. It's not trained to accomplish a task sequentially but instead learns from the inputs and solves it all at once.

Ethical considerations: An awareness of the ethical implications of AI and issues related to privacy, data usage, fairness, misuse and other safety issues.

Foom: Also known as fast takeoff or hard takeoff. The concept that if someone builds an AGI that it might already be too late to save humanity.

Generative adversarial networks, or GANs: A generative AI model composed of two neural networks to generate new data: a generator and a discriminator. The generator creates new content, and the discriminator checks to see if it's authentic.

Generative AI: A content-generating technology that uses AI to create text, video, computer code or images. The AI is fed large amounts of training data, finds patterns to generate its own novel responses, which can sometimes be similar to the source material.

Google Gemini: An AI chatbot by Google that functions similarly to ChatGPT but pulls information from the current web, whereas ChatGPT is limited to data until 2021 and isn't connected to the internet.

Guardrails: Policies and restrictions placed on AI models to ensure data is handled responsibly and that the model doesn't create disturbing content.

Hallucination: An incorrect response from AI. Can include generative AI producing answers that are incorrect but stated with confidence as if correct. The reasons for this aren't entirely known. For example, when asking an AI chatbot, "When did Leonardo da Vinci paint the Mona Lisa?" it may respond with an incorrect statement saying, "Leonardo da Vinci painted the Mona Lisa in 1815," which is 300 years after it was actually painted.

Large language model, or LLM: An AI model trained on mass amounts of text data to understand language and generate novel content in human-like language.

Machine learning, or ML: A component in AI that allows computers to learn and make better predictive outcomes without explicit programming. Can be coupled with training sets to generate new content.

Microsoft Bing: A search engine by Microsoft that can now use the technology powering ChatGPT to give AI-powered search results. It's similar to Google Gemini in being connected to the internet.

Multimodal AI: A type of AI that can process multiple types of inputs, including text, images, videos and speech.

Natural language processing: A branch of AI that uses machine learning and deep learning to give computers the ability to understand human language, often using learning algorithms, statistical models and linguistic rules.

Neural network: A computational model that resembles the human brain's structure and is meant to recognize patterns in data. Consists of interconnected nodes, or neurons, that can recognize patterns and learn over time.

Overfitting: Error in machine learning where it functions too closely to the training data and may only be able to identify specific examples in said data but not new data.

Paperclips: The Paperclip Maximiser theory, coined by philosopher Nick Boström of the University of Oxford, is a hypothetical scenario where an AI system will create as many literal paperclips as possible. In its goal to produce the maximum amount of paperclips, an AI system would hypothetically consume or convert all materials to achieve its goal. This could include dismantling other machinery to produce more paperclips, machinery that could be beneficial to humans. The unintended consequence of this AI system is that it may destroy humanity in its goal to make paperclips.

Parameters: Numerical values that give LLMs structure and behavior, enabling it to make predictions.

Parameters: Numerical values that give LLMs structure and behavior, enabling it to make predictions.

Prompt: The suggestion or question you enter into an AI chatbot to get a response.

Prompt chaining: The ability of AI to use information from previous interactions to color future responses.

Stochastic parrot: An analogy of LLMs that illustrates that the software doesn't have a larger understanding of meaning behind language or the world around it, regardless of how convincing the output sounds. The phrase refers to how a parrot can mimic human words without understanding the meaning behind them.

Style transfer: The ability to adapt the style of one image to the content of another, allowing an AI to interpret the visual attributes of one image and use it on another. For example, taking the self-portrait of Rembrandt and re-creating it in the style of Picasso.

Temperature: Parameters set to control how random a language model's output is. A higher temperature means the model takes more risks.

Text-to-image generation: Creating images based on textual descriptions.

Tokens: Small bits of written text that AI language models process to formulate their responses to your prompts. A token is equivalent to four characters in English, or about three-quarters of a word.

Training data: The datasets used to help AI models learn, including text, images, code or data.

Transformer model: A neural network architecture and deep learning model that learns context by tracking relationships in data, like in sentences or parts of images. So, instead of analyzing a sentence one word at a time, it can look at the whole sentence and understand the context.

Turing test: Named after famed mathematician and computer scientist Alan Turing, it tests a machine's ability to behave like a human. The machine passes if a human can't distinguish the machine's response from another human.

Weak AI, aka narrow AI: AI that's focused on a particular task and can't learn beyond its skill set. Most of today's AI is weak AI.

Zero-shot learning: A test in which a model must complete a task without being given the requisite training data. An example would be recognizing a lion while only being trained on tigers.

Conclusion

AI will have a massive impact on society and the reality is that we are in the early stages of understanding the short and long term implications. However, it is fairly certain that the insurance industry will play a pivotal role in helping businesses and governments manage the risk.

We hope that this “Handbook” is a helpful tool to begin to understand the risks and appreciate one of the early solutions.

For questions or additional information, please contact:

Jonathan Legge | Senior Managing Director
Private Equity and Transactional Liability
e: jlegge@one80.com | p: 203-315-3499 (x5908)

Artificial Intelligence Warranty Coverage

For more information visit:
www.One80.com

Experts predict that third-party AI tools will increase workforce productivity by as much as 40% and add $14 trillion to the global economy by 2035. Despite the benefits of this rapidly evolving technology, it is estimated that third-party AI tools are responsible for over 55% of AI-related failures in organizations.

With that, One80 is pleased to offer *AI Product Warranty Coverage*, protecting enterprises and third-party AI vendors against losses related to unreliable AI models.

In partnership with Armilla AI, the first of its kind coverage verifies open source and proprietary AI models. This allows companies to deploy AI solutions safely and provides enterprises with confidence in the technology procured from third-party AI vendors.

Targeted Industry Classes

Third-party AI vendors

Enterprises which are actively leveraging AI including the following industries:

- ✓ Healthcare companies
- ✓ Retail industry
- ✓ Supply chain exposures
- ✓ Education sector
- ✓ Media and entertainment
- ✓ Travel industry
- ✓ Legal practices

Product Features

- Underwritten by Armilla and backed by top reinsurers.
- Data verification process assesses data quality, qualitative and quantitative criteria, key performance indicators, process flows and overall performance.
- AI models assessed based on requirements set out in legislations such as EU AI Act, and other US federal and state regulations.

Available Coverages

- ✓ Third party guarantee by Armilla
- ✓ Guarantees performance, fairness and robustness of vendor's AI powered products
- ✓ Protects enterprises if performance metrics are not met once model is in operation
- ✓ Reduces uncertainty related to an investment related to an AI model

Contact Us:

Jonathan Legge
Senior Managing Director
e: jlegge@one80.com
p: 203-315-3499 (ext. 5908)

One80 Intermediaries is a privately held firm with offices throughout the US and Canada. As a leading insurance wholesaler and program manager, the company offers placement services and binding authority for property and casualty, life, travel/accident and health, benefits, affinity and administrative services and warranty business. One80 serves commercial companies, non-profits, public entities, individuals and associations and unions, and has access to all major insurance markets in the US, Canada, Europe and Asia. One80 has offices in more than 55 locations in the US and Canada including Boston, New York City, Chicago, Houston, Philadelphia, San Diego, Seattle, Toronto and Montreal.

V052024