

**Textmetrics**

# Address the world

For everyone who writes with people and goals in mind

Safeguard your use of Generative AI with **Textmetrics** .

## Summary

As companies begin to explore the use of AI in their business strategies, the discussion inevitably turns to the several risks inherent to the broad adoption of AI.

Textmetrics proposes to mitigate many of those risks by integrating functionality that will make AI the reliable partner in business that it can be.

We propose an integration model that will separate the business logic, UI and presentation layers from one another, giving reliable structure to the business logic results and predictability to how those results are presented to the UI and, by extension, to its end users. This "separation of concerns" illustrates the technology relationship proposed in this document.

As fully integrated functionality, Textmetrics offers a mature feature set to clients and their end users; a powerful, versatile content creation assistant and a safety net when using nascent, powerful technology.

## Reference Points

- [Risks](#)
- [Personas & Scenarios](#)
- [Conclusion](#)
- [Further Reading](#)

# Risks

As discussed in our powerpoint presentation of April, 2, 2024, there are six risks inherent to the use of AI in a business setting:

- **Risk 1** – Users rely on the veracity of answers generated by the AI and do not feel the need to check the answers independently. Because AI can be prone to fabricating responses, quality and reputation may suffer if this information is utilized without first being vetted and checked for accuracy.
- **Risk 2** – For data to be usable, it must be structured, accessible and accurate. Therefore, there is a need to standardize Processes, Data utilization, Technology and Quality Control. In order to make the most of AI, one must also use tools that support data privacy and confidentiality.
- **Risk 3** – Implicit bias present in the model and output will not be addressed by generative AI technology. This bias is present due to programming choices and the data used in training. In order to minimize the use of biased material, the output should be reviewed by technology that can identify the bias, determine the quality of the output and highlight risks.
- **Risk 4** - The underlying model of the AI changes over time altering or degrading the quality of the output. The LLM may demonstrate a bias in the future that was not present at the time the LLM was originally adopted due to updates to the data upon which the LLM is trained. There is the risk that the quality of output changes.
- **Risk 5** – Much of the data with which the LLM is trained is proprietary intellectual property (IP), protected by copyright and may have been used without permission. Therefore, the LLM was trained based upon data which is not the property of the owner of the LLM. The training data used should be reviewed and copyrighted material deleted before being processed by the LLM, unless permission to use that material is granted by its owner.
- **Risk 6** – Give users a defined set of tools instead of the whole toolbox. By giving the users defined usage of an LLM (e.g.: predefined prompts), the risk of misuse is lowered, while at the same time the usability of the system increases dramatically. Defined usage will lower the possible risks while at the same time increasing the usability and adoption of AI for specific business purposes.

(Adapted from article "[Gartner Survey Shows Generative AI Has Become an Emerging Risk for Enterprises](#)")

AI-generated content has transformed content creation. Because of the rising challenge of plausible yet inaccurate or fabricated information, together with potential copyright issues related to GPT's training data and potential implicit bias, business users will be wise to look for a robust approach to governance. We believe that Textmetrics is the governance tool of choice as it easily addresses accuracy concerns and the risks associated with copyright issues.

Textmetrics focuses on these primary risks:

- Apparent Plausibility vs. Accuracy
- Copyright Issues in GPT's Training Data
- Guarding Against the Use of AI in a Prohibited Setting
- Risks to the Quality of Output, Now and in the Future

## Personas and Scenarios

To illustrate the potential uses of Textmetrics as an AI governance tool, we have created a fictitious AI, Penrose Technology's Penrose AI, and a fictitious corporate client of the AI, Temporal Tech Solutions, creator of time keeping solutions for companies all over the world, and Iris Wright who is an HR specialist at Temporal Tech. These use cases will show some of the potential risks that Iris encounters as she uses Penrose in the course of her work.

### Personas

#### *Penrose AI Technology*



*"Making the impossible, possible."*

**Product:** Penrose AI

Assist individuals and organizations across various industries in tasks such as content generation, language translation, information retrieval, and more.

**Role:** Platform Integrated AI

#### *Temporal Tech Solutions*



*"Your time, our craft."*

**Product:** TimeCraft Suite

Provides time tracking for global teams, integrating with project management tools, spreadsheets, and ERPs.

**Role:** Platform Client

## *Iris Wright, HR Specialist at Temporal Tech Solutions*



Responds to human resources concerns presented by employees of TemporalTech.

**Task:** Writes new employee policies, procedures and guidelines for employees using a content creation platform, Penrose and Textmetrics.

**Role:** Platform User

### **Scenarios**

#### ***Risk 1: Misinformation***

**Issue:** Iris uses Penrose to help her write a policy based on the new insurance guidelines required by an employee's state of residence. The AI misinforms her, referencing a statute from another state.

**Mitigation:** Textmetrics' state-of-the-art algorithms verify content accuracy and credibility, reducing risk of misinformation.

- Ensure quality and reliability of content, enhancing its trustworthiness and minimizing dissemination of inaccurate or misleading information.
- Analyze text, identify potential errors or inconsistencies, and provide valuable insights to improve the overall accuracy, credibility and quality of the content.

**Resolution:** Textmetrics alerts Iris to the incorrect statute and Iris refines her prompt to include only those statutes from the state in question.

#### ***Risk 2: Unstructured Data***

**Issue:** Iris uses Penrose to ideate an explanation of how the Family Leave Act works in accordance with company guidelines and the new insurance policy. Penrose provides her with irrelevant policy pricing data and unorganized, vague data.

**Mitigation:** Iris refers to her AI prompt guide that she has developed as she's worked with Penrose for guidance to better structure her prompt to generate this content.

**Resolution:** Iris fine-tunes her prompt in order to generate more relevant content and data from Penrose.

### ***Risk 3: Poor Quality Output***

**Issue:** Iris is using Penrose AI to ideate a policy on racial discrimination in the workplace and Penrose generates content using the word “Oriental” to refer to persons of Asian heritage.

**Mitigation:** Textmetrics reviews the output for bias, determines the quality of the output and highlight risks.

**Resolution:** Textmetrics’ bias filter warns Iris of the use of the biased term and makes suggestions of terms or phrases that would be more appropriate to corporate communications.

### ***Risk 4: Output Quality Diminishes Over Time***

**Issue:** After a recent update to Penrose’s data pile, Iris prompts the AI to ideate an article for the company newsletter about an intramural sports league that the company is sponsoring in the spring. Penrose is using terms in the article to describe the teams that are inaccurate and misleading.

**Mitigation:** Textmetrics reviews content for readability, credibility, implicit bias, style, sentiment and gender target. Textmetrics is rules-based software, offering suggestions through the lenses with which it has been trained. These rules allow users to “future proof” content against changes to an LLM that may cause it to exhibit biases down the road that were not present at the time of adoption.

**Resolution:** Textmetrics flags the terms and alerts Iris, offering suggestions that are appropriate to company communications per the style guide ruleset with which it was programmed.

### ***Risk 5: IP Used Without Permission***

**Issue:** Iris is writing an article for the upcoming company newsletter about market trends in a new line of business for her firm. Iris is unaware that some of the market information that she is referencing in the article was published by a business consulting firm with which they do not have a relationship.

**Mitigation:** Textmetrics has the potential to facilitate compliance with copyright regulations by ensuring transparency and adherence to permissions for content used in AI training data. By incorporating features that promote transparency and respect for intellectual property rights, Textmetrics will help organizations navigate the complexities of copyright laws when utilizing AI-generated content.

**Resolution:** Iris cites another source for data that she will use in her article.

## ***Risk 6: AI Use in Prohibited Setting***

**Issue:** Iris uses AI to ideate a report packet about a new division of the company and inadvertently includes proprietary market research that she was asked to exclude. These reports will be used in an upcoming Board meeting and are then to be transmitted to the Investor Relations team.

**Mitigation:** Textmetrics recognizes that proprietary information is being included in the report requests and suggests to Iris that AI ideation is not appropriate in this instance due to the sensitivity of the report.

**Resolution:** Iris reaches out to the Marketing team for a sanitized version of their information for inclusion in the report.

## **Conclusion**

Feel free to contact us with questions or a request for demonstration:

Kyle Schweighauser [kms@extanto.com](mailto:kms@extanto.com)

Ben Berry [ben@extanto.com](mailto:ben@extanto.com)

## **Further Reading**

### **Extanto White Paper About AI Risk**

For business to adopt AI, it must trust AI. To ensure ethical and reliable content creation, a ruleset is imperative for governing GPT-generated content. Textmetrics is a powerful governance tool, addressing bias and accuracy concerns and mitigating the risks associated with potential copyright issues in GPT's training data...( [read more](#) )

### **Integrating AI and Textmetrics**

To improve the quality and consistency of AI's output, we propose an integration model that is commonly used to implement user interfaces...( [read more](#) )

### **Limitations of Tonality with GPT**

We've all had the experience in which something that we've written has been misinterpreted by the person to whom we wrote it. ( [read more](#) )

### **Domain Expertise: AI vs Human**

Domain Expertise is the deep understanding of a specific industry or field of study born of the lived experience of someone immersed in that industry or field for an extensive period of time. The idea of Domain Expertise is key to unlocking a better understanding of the value that Textmetrics brings to the AI transaction. ( [read more](#) )

## **The Value of a Smart AI Writing Assistant**

Discover how the Textmetrics Smart AI writing assistant revolutionizes content generation. Taking advantage of the latest AI technologies, we ensure high quality, unique and correctly styled content. ( [read more](#) )

## **Article From Gartner Regarding AI as an Emerging Risk for Enterprises**

Survey of 249 Senior Enterprise Risk Executives Reveals Top 5 Emerging Risks in the Second Quarter of 2023...( [read more](#) )