Progress® Semaphore™

Explainable and Trustworthy AI – How Can a Business Trust an Algorithm?

WHITEPAPER

Artificial intelligence (AI) is not a new trend, though more and more we're seeing its applications in our everyday lives and in the enterprise. AI is not only a tool for spurring innovation; it has become a crucial part of the organization's technological landscape. It offers new and innovative opportunities for businesses, enabling them to gain better insights, optimize processes, drive innovation, analyze information and more. The world is changing, and we see AI technologies, like generative AI, taking a leading role within this new digital environment. AI is not just another technology that we need to get used to, but a technology that we should embrace responsibly to flourish within our new digital landscape.

Many enterprises have already embarked on their AI journeys. As AI systems become more prevalent and influential, promoting trust and transparency in decision-making processes is paramount. Enter explainable and trustworthy AI. With semantic technologies, like Progress® Semaphore™, companies can achieve more transparent, contextual and reliable AI outcomes.

Business Challenges Associated with AI Technologies

Today's AI applications and the concept of Machine Learning (ML) create a false expectation that we've reached **The Singularity**—when machines act as fully functioning autonomous units that perceive, learn, decide and act on their own. In their current form, the capabilities and effectiveness of AI systems are limited by several factors:

- Experienced humans, also known as Subject Matter Experts (SMEs), are the source of all business knowledge.
- AI machines need business knowledge in context to reason and perform the requested data operations.
- Machines derive context from knowledge models curated by SMEs and/or unbiased training sets of data and/or documents.
- Unbiased training sets are difficult and costly to assemble, and the machine training is performed by specialist data scientists, not SMEs who understand the data—disconnecting business intellect and the meaning of data.
- Business users don't trust AI machine results. AI systems are unable to explain, to a user's satisfaction, how decisions are made, which actions are taken and what data is used in processing.

In simple terms, explainable and trustworthy AI encompasses AI technology that is transparent and reliable in its operations so that users can understand and trust its business decisions. Today, AI does not live up to the trustworthy and explainable task as it:

- Uses complex mathematical algorithms over vast amounts of data
- Applies algorithms by a specialist data science expert, not a SME
- Delivers noisy results that are difficult to explain and understand
- Fails unexpectedly and can be difficult to debug

While AI can help effectively enhance and contribute to many real-world use cases, it does come with its fair share of downsides. The inability to explain can hurt us when we use AI for high-stakes applications, such as biotech, public health, agriculture, financial services and defense/intelligence. Service companies, technology companies and government agencies are wrestling with these issues and investing time and resources to identify standards and create and implement systems that are explainable, reliable and trusted.

Depending upon the seriousness of the application and the consequences of errors, low-level explainability might be enough. In more serious applications, such as financial services and clinical diagnosis, an exceedingly high degree of explainability is required and when not provided, businesses could face large fines and a damaged reputation.

Leveraging Explainable and Trustworthy AI

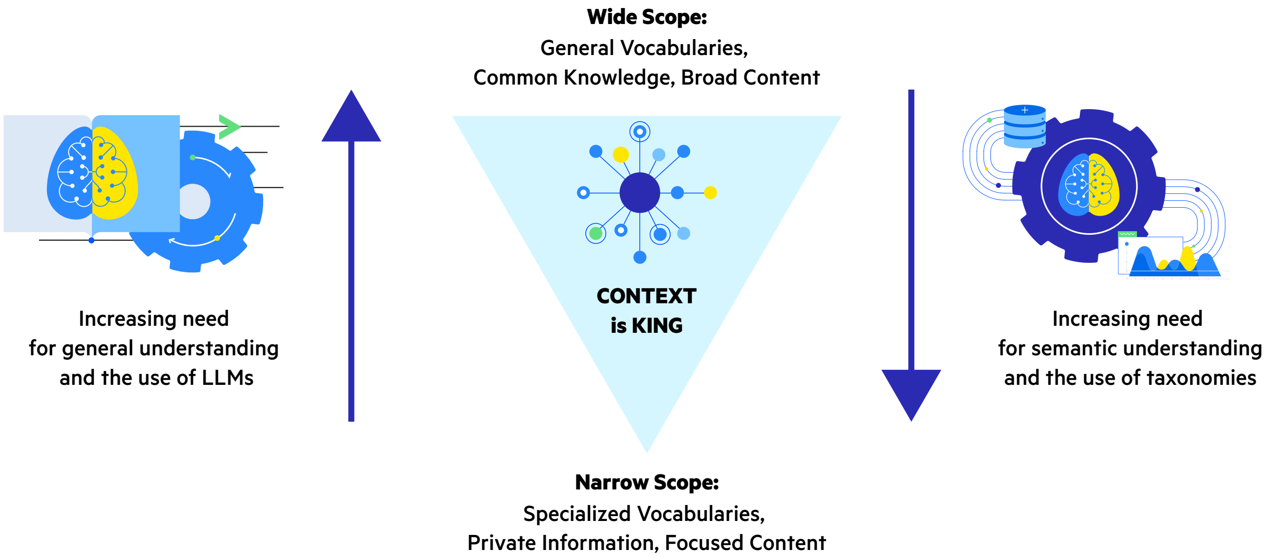
The answer to “how much explainability or trustworthiness is enough?” depends on the situation. For example, if we’re predicting the outcome of the winner of the Best Actor award from the Oscars, the need for transparency is far less than predicting and recommending a treatment plan for a cancer patient.

Mainstream AI is pervasive—we see it online with targeted ads and product recommendations. Facial recognition has transformed our phones, giving us access to nearly any app. Spell checks have evolved by offering prediction options for your next word. It’s also used in image processing and gaming applications. In these scenarios, data is nearly free and unlimited, action takes priority over learning and there is little to no cost of failure. The need for explainability in these situations is low.

While not perfect, AI can be transformational for businesses. It can increase work efficiency, work with high accuracy, reduce the cost of training and operations and improve processes. However, AI can work better as an addition to human skills, not a replacement for them. Designers of AI systems must guide the development and application of AI to fit within social norms and values. AI must be leveraged in a way that won't amplify problems or lead to a bad reputation and will deliver accurate results in business use cases.

Context: The Link to Explainability and Trustworthiness

AI can be very powerful in broad areas that require a general understanding of the world, however, whenever a broad focus occurs, there's a risk of missed context and inaccurate results. As context gets more focused (and vocabulary gets more specialized), the need for semantic understanding increases.



Wide Scope:
General Vocabularies,
Common Knowledge, Broad Content

Increasing need
for general understanding
and the use of LLMs

CONTEXT
is KING

Increasing need
for semantic understanding
and the use of taxonomies

Narrow Scope:
Specialized Vocabularies,
Private Information, Focused Content

Human language is nuanced and ambiguous. For example, the phrase "We saw her duck" might mean that we saw a woman bend to avoid something (duck as a verb) or we may have seen her pet duck (noun). Without context, it's impossible to determine.

Conveying ideas and reacting appropriately to them comes easily to humans thanks to the richness of their shared language, the common understanding of how the world works and an implicit awareness of everyday situations. When humans talk with humans, they increase their conversational bandwidth using context.

Unfortunately, humans are less successful when they attempt to convey ideas to computers. In traditional interactive computing, computers are unable to leverage context in the human-computer dialogue. Computers are unable to handle arbitrary inputs and fail when unexpected responses are given; for example, if the human user responds to a “Yes” or “No” question with “I dunno,” the system will likely provide an incorrect result.

AI systems must be:

- **Robust:** Context provides predictive accuracy and situational awareness. As things change, context will be the key to understanding that change.

- **Trustworthy:** Results that are fair, reliable and explainable

  - Fairness – Through qualification of past and current data and verification that it’s free of bias and unskewed by discrimination, fair results can be attained.

  - Explainable – By knowing how data is collected, where it’s from, when and how it’s collected and the processes and features used in the training process, bias can be identified and eliminated.

  - Reliable – When underlying information is explainable, it prevents data manipulation, and all information related to the task can be incorporated into the process, bringing data silos together.

AI technologies need to implement context and be able to refine judgments/decisions when new information is presented. Without context, AI is narrowly focused, makes subpar predictions and has limited transparency. AI that is powered by semantics provides context to the human-machine interaction.

Semantic AI Technologies Provide Context

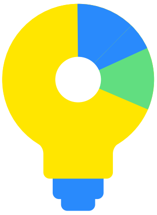
Semantic AI technologies provide a structured framework for capturing and representing the semantics of data. By encoding rich semantic relationships between entities, attributes and concepts, these technologies enable more precise and contextually relevant information retrieval.

Semantic AI is more than “another machine learning algorithm,” it combines several AI technologies including Machine Learning (ML), Natural Language Processing (NLP) and semantic reasoning as a collaborative interplay between humans and machines to drive analytics, automation and insights. This approach allows machines, as well as people, to understand, share and draw conclusions from data—structured, semi-structured or unstructured—whether internal or external to an organization. From a user perspective, semantics provide far more intelligent, capable, relevant and contextual interaction than with traditional information technologies.

Different from traditional information processing methods, Semantic AI encodes meaning and context separately from data, content files and application code. This allows analytics, insight and information automation projects to be quickly implemented and deliver high data veracity. Meaning and context are controlled and moderated by SMEs who capture domain knowledge in knowledge models. The Semantic AI engine uses the model to deliver this as high-quality metadata which is "reasoned" through data.

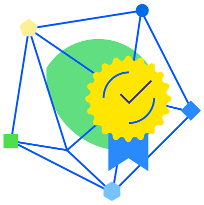
Semantic AI provides an abstraction layer that enables the bridging and interconnection of data, content and processes without requiring human intervention every time there is a change. With semantics, adding, changing and implementing new relationships or interconnecting programs in a new way can be as simple as changing the external model the programs share. This is different than traditional methods, which require meanings and relationships to be predefined and "hard-wired" into data formats and application code at design time.

Semantically enriched data supports knowledge-centric architectures, enabling companies to use enterprise knowledge more effectively through data quality enhancement, data enrichment, data governance and knowledge management practices. These are critical for enabling AI models to deliver quality results.



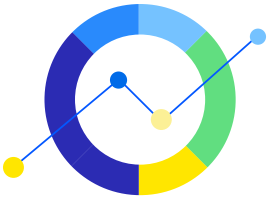
Knowledge Governance

Semantic AI is moderated and instructed by business users. SMEs are responsible for the management, development and maintenance of knowledge models over their lifetime. The models represent relevant knowledge in the language and vocabulary used by the business to provide qualified contextual data.



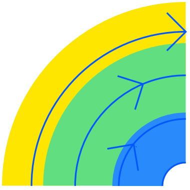
Data Quality

Semantic metadata—data that is harmonized, enriched and extracted—provides a holistic view of all enterprise information. This includes data that is structured and unstructured, as well as internal and external to the organization. It improves data quality, reduces noise and results in a higher precision of prediction, which is important, as the quality of information affects outcomes.



Auditable Outcomes

Semantic AI provides a transparent approach for harmonizing information differences. As the semantic platform evaluates the information and makes decisions, the knowledge model and metadata provide a clear and explicit picture of the information used and actions taken in the decision-making process.



Repeatable Outcomes

Semantic AI provides precise, complete and consistent results. By leveraging a knowledge model with rule-based classification and fact extraction, information is processed consistently, and decisions and outcomes are repeatable, transparent and fact-based.

Semantics are a key component in AI as they transcend industry, organization and use case. Semantic AI enables safe and responsible applications and systems that are trustworthy, transparent and aligned with societal norms and values.

Context Enables Responsible, Trustworthy and Transparent AI Results

Unlike traditional software systems where humans define the logic of computation, AI systems are trained with vast amounts of data, which may contain human bias, so we don't really know how these systems make decisions. As these systems become more complex, the expectations for global AI systems must also go beyond performance or accuracy on curated data—they must be responsible and focus on ethics, fairness and explainability.

- **Ethics:** Multiple stakeholder positions must be evaluated and considered. Often what is beneficial to the algorithm designer is not beneficial to others.

- **Fairness:** Systematic bias in data must be eliminated so one group of people will not benefit more than another.

- **Explainability:** The black-box nature of current AI systems makes understanding solutions difficult; explaining why a decision was made is as important as the decision itself.

Context derived from Semantic AI systems can facilitate responsible AI. Through metadata, which is reasoned through data, it provides robust and trustworthy results. It addresses the issues of fairness and reliability by eliminating bias and providing explainable and transparent outcomes.

Yet even if AI applications and systems have context and incorporate curated as well as non-curated data, can these systems provide explainable and transparent results? Semantic AI Technology combines semantics with technological advances to enable organizations to create fair, transparent and trustworthy AI applications and systems.

The Future – How Do We Make AI Safe?

As AI becomes more ingrained in our lives, explainable AI becomes even more important. AI is designed to learn from interactions with its surroundings and alter behavior, decisions and outcomes as appropriate. While it has the potential to provide incredible benefits, it's not hard to imagine how it could go wrong.

While researchers can learn from inappropriate or unintended AI results, the lack of transparency makes it difficult to determine how or why AI has taken a particular action. And as AI becomes more complex, this process will become more challenging—and more critical.

No one level of transparency can be applied to all applications, systems and use cases. While it's difficult to standardize algorithms or explainable approaches, it is possible to standardize levels of transparency based on consequences. The more harmful the consequence the greater the level of explainability and transparency.

Organizations must apply governance and oversight to AI systems:

- **Take a heuristic approach to AI.** Use formal and independent risk assessment methods—such as committee reviews, checklists, EU guidelines—in the model development phase.

- **Verify the accuracy of AI results.** Are your predictions reasonable? Does it predict what you want?

- **Vet your model.** Know the logic that is used and use a simpler model that's interpretable and documented.

- **Leverage Semantic AI technologies.** Knowledge models and rule-based auto-classification provide context through precise and consistent metadata. Context drives explainable, traceable and consistent outcomes.

- **Incorporate knowledge graphs.** These can be used to identify relationships found in the data.

- **Keep a human in the loop.** For high-stakes decisions, insist on explainable results that are accurate, complete and faithful. Have subject matter experts validate AI outcomes and take control when required.

As we go forward, human values will increasingly impact AI systems and context will be a necessary component to make AI responsible, reliable and safe. To move from the questions we have today to a solid foundation for future AI applications, semantic AI technologies must be a key component.

Learn how Semaphore can help you leverage semantic knowledge for trustworthy generative AI. **Contact us today.**

About Progress

Progress (Nasdaq: PRGS) empowers organizations to achieve transformational success in the face of disruptive change. Our software enables our customers to develop, deploy and manage responsible AI-powered applications and experiences with agility and ease. Customers get a trusted provider in Progress, with the products, expertise and vision they need to succeed. Over 4 million developers and technologists at hundreds of thousands of enterprises depend on Progress. Learn more at www.progress.com

Worldwide Headquarters

Progress Software Corporation
15 Wayside Rd, Suite 400, Burlington, MA 01803, USA
Tel: +1-800-477-6473

facebook.com/progresssw
twitter.com/progresssw
youtube.com/progresssw
linkedin.com/company/progress-software
progress_sw_