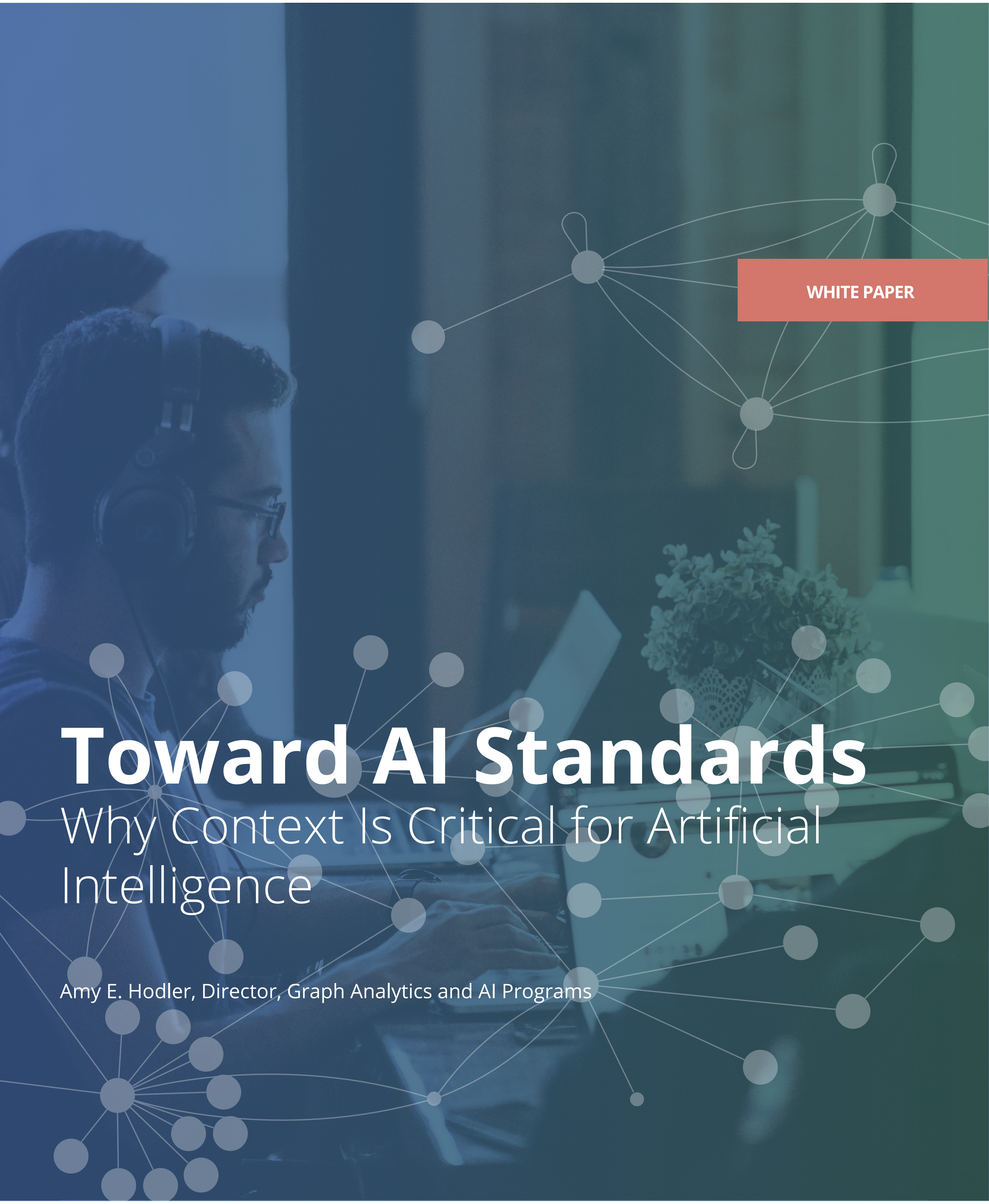
The #1 Platform for Connected Data



WHITE PAPER

Toward AI Standards

Why Context Is Critical for Artificial Intelligence

Amy E. Hodler, Director, Graph Analytics and AI Programs

White Paper

TABLE OF CONTENTS

Toward AI Standards

Why Context Is Critical for Artificial Intelligence

Amy E. Hodler, Director, Graph Analytics and AI Programs

Introduction

The potential power of artificial intelligence is expansive and will be used in ways we cannot yet imagine. Despite, and perhaps because of this, we have a duty to guide its development and application in ways that facilitate innovation and fair competition, public trust and confidence, while incorporating the appropriate protections.

Context must be incorporated into AI to ensure that we apply these technologies in ways that do not violate our societal and economic principles. AI must be guided by not just technical standards but ethical standards, and context is the key.

Graph Technology & Creating Context for AI

For artificial intelligence (AI) to be more situationally appropriate and "learn" in a way that leverages adjacency to understand and refine outputs, it needs to be underpinned by context. Context is all of the peripheral information relevant to that specific AI.

AI standards that don't explicitly include contextual information result in subpar outcomes as solution providers leave out this adjacent information. The result is more narrowly focused and rigid AI, uninterpretable predictions and less accountability.

Graph technologies are a state-of-the-art, purpose-built method for adding and leveraging context from data. Proven repeatedly in deployments worldwide, graph technology is a powerful foundation for AI.

Artificial Intelligence & Contextual Information

AI today is effective for specific, well-defined tasks but struggles with ambiguous ones. Humans deal with ambiguities by using context to figure out what's important in a situation and then also extend that learning to understanding new situations.

For example, if we are making travel plans, our decisions vary significantly depending on whether the trip is for work or pleasure. Once we learn a complicated or nuanced task, like driving a manual transmission, we easily apply that to other scenarios, such as other vehicles. We are masters of abstraction and recycling lessons learned.

For artificial intelligence to make human-like decisions that are more situationally appropriate, it needs to incorporate context – all of the adjacent information. Context-driven AI also helps ensure the explainability and transparency of any given decision, since human overseers can better map and visualize the decision path.

Without context, AI requires exhaustive training, strictly prescriptive rules and specific applications. Moreover, since it's impossible to anticipate every possible situation, we often find AI solutions wanting when new situations arise. Sometimes outcomes are even harmful such as biased recommendations or harmful interpretations.

For example, Microsoft's Twitter bot, Tay, learned from Twitter users how to respond to tweets. After interacting with real social media users, however, Tay learned offensive language and racial slurs. Another example is Amazon's AI-powered recruiting tool, which was shut down after showing bias against women candidates (as reported by Reuters).

In both of these cases, the machine learning models were trained on existing data that lacked the appropriate context. Tay's model drew more often from the loudest and most outrageous opinions, reinforcing those as the norm. Amazon's recruiting tool amplified and even codified discriminatory practices because its existing dataset was too narrow.

And how do we identify whether an AI-based solution is suboptimal, incorrect or bad? Do we wait until something terrible happens?

Knowing if an AI project has gone off course requires a larger frame of reference to identify how millions of data points and procedures come together. If we fail to evaluate AI outcomes within a larger context, we risk accelerating unintended outcomes as data and technologies become more complex.

AI explainability is also critical for accountability. Particularly in nuanced situations, such as creditworthiness or criminal sentencing, unchecked AI runs the potential of putting entire groups at a disadvantage.

Context-driven AI helps us understand and explain the factors and pathways of logic processing so organizations better understand AI decisions.

Context Makes AI More Robust

Context-driven AI helps us understand and explain the factors and pathways of logic processing so organizations better understand AI decisions.

Better Predictions

One of the greatest challenges in training artificial intelligence models is gathering enough relevant data. And yet, current methods don't incorporate existing relationships within data, essentially throwing out predictive information. Using context adds relevant information and results in better predictions – all with only the data we already have.

According to Stratistics MRC, the global fraud detection and prevention market was valued at $17.5 billion in 2017 and is expected to grow to $120 billion by 2026, and over 48,000 U.S. patents for graph fraud/anomaly detection have been issued in the last 10 years.

Many financial services companies are tapping into graph technology to reveal predictive patterns, find unusual behavior and score influential entities. All of this contextual information is then loaded into their machine learning models.

Even beyond the obvious fit in financial services, people are using graph algorithms in various industries to create (engineer) more predictive "features" that train AI models for higher accuracy and precision.

> Today's data is highly connected and has uneven concentrations, which basic statistics and averages completely miss.

For example, the Association for the Advancement of Artificial Intelligence used graph algorithms to detect clusters of interactions between doctors and pharmacies to improve opioid fraud predictions as described in "Graph Analysis for Detecting Fraud, Waste, and Abuse in Health-Care Data." Analyzing the graph revealed multiple anomalies including pharmacies with the most narcotics revenue from a small patient population, unusually high volume between certain doctors and pharmacies, and patients visiting multiple doctors to obtain narcotics.

Today's data is highly connected and has uneven concentrations, which basic statistics and averages completely miss. Graph algorithms are specifically developed to leverage the topology of data through connections: find communities, uncover influential components and infer patterns and structure.

Incorporating the predictive elements of context and relationships into machine learning greatly increases model accuracy and precision – with the data we already have.

Situational Flexibility

Situational awareness and appropriateness is another concern for AI where context-based learning and action are critical.

For example, consider how we want an age-appropriate chatbot sensing and responding differently in an interaction with a 7-year-old versus a 30-year-old. In fact, an investigation highlighted an egregious example: a mental health chatbot created for use by children was unable to understand a child explicitly reporting underage sexual abuse as reported in the BBC news exposé, "Child advice chatbots fail to spot sexual abuse."

AI-based systems need to be flexible, which includes designing AI in a way that views user interaction as critical to the design and implementation of autonomous decision-making systems. This user-centric thinking might have helped prevent the recent loss of two Boeing 737 MAX aircraft, which investigations revealed were partly due to the failure to incorporate pilot behavior into automated systems.

Contextual information also helps an AI solution flex within new situations that it is untrained for, reducing failures and equipping it with new data or unexpected scenarios.

For example, a semi-autonomous car might be programmed to slow down in rainy weather, but we would also want it to expand its AI application to incorporate contextual information such as falling temperature and an approaching bridge. Or especially difficult, apply learning to break for objects in the road such as a dog but not slam on the breaks for a paper bag or flock of birds.

Finally, when AI solutions are based on contextually aware and dynamic backends they are more broadly applicable. In turn, they will spur new innovations and broaden competition.

Reliability

In order for AI solutions to be reliable and fair, we need to know what data was used to train our models and why. Unfortunately, this isn't as straightforward as we might think. If we consider a large cloud service provider or a company like Facebook with an enormous amount of data, it's difficult to know what exact data was used to inform its algorithms. (Let alone which data may have changed!)

Graph technology adds the required context for this level of explainability.

For example, graph technology is often used for data lineage to meet data compliance regulations such as GDPR or the California Consumer Privacy Act (CCPA). A data lineage approach is also used on NASA's knowledge graph to find past "lessons learned" that are applicable to new missions. When we store data as a graph, it's easier to track how that data is changed, where data is used and who used what data.

Understanding and monitoring data lineage also guards against the manipulation of input data.

For example, corporate fraud research has shown that when the significance of input data is common knowledge, people will game the system to avoid detection. Imagine a utility system or network infrastructure where we were confident in our monitoring software, but could no longer rely on the input data. The whole system would become immediately untrustworthy.

Finally, context-driven AI systems avoid excessive reliance on any one point of correlated data.

For example, organizations are beginning to apply AI to automate complex business dependencies in areas such as data centers, batch manufacturing and process operations. With contextual coordination, they avoid the trap of noisy, non-causal information and use root-cause analysis to maximize future efficiency. Contextual information helps us identify the root cause of a problem as opposed to just treating a symptom.

Fairness

Understanding the context of our data also reveals the potential biases inherent in existing data as well as how we collect new data and train our models.

For instance, existing data may be biased by the fact that it was only collected for one gender. Or perhaps an AI's human language interactions were trained on a narrow age or accent range. Graphs ensure situational/contextual fairness by bringing context to the forefront of AI solutions.

Higher arrest rates for some demographic populations become embedded in prosecution data. When historical input data is used for predictive policing, it causes a vicious cycle of increased arrests and policing.

The Royal Statistical Society published an Oakland, CA simulation analysis of a machine learning approach often used for predictive policing and found "that rather than correcting for the apparent biases in the police data, the model reinforces these biases."

Fair use of our personal data is an important part of AI-based systems. Contextual data can assist in privacy efforts because relationships are extremely predictive of behavior. This means we can learn from peripheral information and collect less information that is personally identifiable.

In the book *Connected*, James Fowler describes studies that have shown that even with little or no information about an individual, we can predict a behavior such as smoking based on the behavior of friends, or even friends of friends.

Trust & Explainability

Training a machine learning model is mostly done on existing data, but not all situations can be accounted for ahead of time. This means we'll never be completely sure of an AI reaction to a novel condition until it occurs.

Fair use of our personal data is an important part of AI-based systems.

To increase public trustworthiness of AI, predictions must be more easily interpretable by experts and explainable to the public.

Consequently, AI deployments need to dynamically integrate contextual information. For example, researchers have developed an application that predicts the legal meaning of "violation" based on past cases. However, legal concepts change over time and need to be applied to new situations. Because graphs capture and store relationships naturally, they help AI solutions adjust faster to unexpected outcomes and new situations.

To increase public trustworthiness of AI, predictions must be more easily interpretable by experts and explainable to the public. Without this, people will reject recommendations that are counterintuitive.

Graph technologies offer a more human-friendly way to evaluate connected data, similar to drawing circles to represent entities and lines to show how they are connected on a whiteboard.

There are also new ways to learn based on contextual information. For example, companies like eBay are mapping the potential pathways a person might take from one purchase to another in order to recommend another selection.

In the example of a music download, there's a lot of context around a person and their selection: the artist, album, publishing decade, music genre, etc.

A paper titled "Explainable Reasoning over Knowledge Graphs for Recommendation" details how to use graph technology and machine learning to predict the path (from song to artist, album, genre or decade, etc.) that a person would likely take to get to their next song purchase.

We can use graphs to relay the complex data of people, songs, albums and their relationships into practical machine learning measures while retaining the various potential paths from one purchase to another. When we combine the likelihood of different paths with context, we better understand such AI-powered recommendations and decision-making.

Oversight opportunities also benefit from increased AI explainability. When homeowners insurance increases for an individual, it's frustrating; when rates increase for an entire demographic, it's discrimination. Without efficient AI explainability, it will take regulators considerable more effort to determine the root cause of such discrimination.

The Promise of Graph-Native Learning

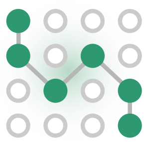
We're excited by recent work that appears to offer a leap forward for machine learning explainability. Particularly promising is the idea of graph-native learning, which involves computing machine learning tasks within a native graph structure itself to make use of its natural context.

Implementing artificial intelligence in a way that is underpinned by connected data as suggested above results in AI solutions that are much more generally applicable.

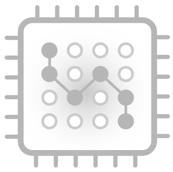
However, perhaps even more significant will be the extreme transparency afforded by this approach: Graph-native learning enables us to input connected data, learn while keeping data connected and then output AI outcomes in the same graph format, which helps those accountable to accurately interpret results.

The Next Major Advancement in AI Will Be Graph-Native Learning

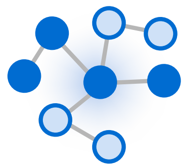
Implements machine learning in a graph environment, creating AI solutions that are more accurate, flexible and trustworthy



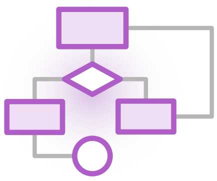
Input data as a graph



Learns while preserving transient states



Output as a graph



Track and validate AI decision paths

More accurate with less data and training

Graph-native learning moves AI from the rigid black boxes of today to extremely flexible and transparent models

In addition, the intermediate states of learning become uniquely observable in that same connected format, which means experts track and validate an AI's decision paths.

With the advancement of graph-native learning, we interpret and explain how an AI comes to a particular conclusion – which is a far cry from the black-box approach used today.

Graph technologies are inherently cross-sector, with the Neo4j Graph Platform used globally by:

- 20 of the top 25 financial services firms
- 7 of the top 10 software firms
- 3 of the top 5 logistics firms
- 7 of the top 10 retailers
- 3 of the top 5 airlines
- 4 of the top 5 telecommunications firms
- And 3 of the top 5 hospitality companies

Conclusion

Although any new standards endeavor is difficult and imperfect, organizations around the world are making progress in encouraging reliable, robust and trustworthy systems that use AI technologies.

Furthermore, when considering such a broad and evolving area of technology, widely applicable standards and tools – such as integrating context – should be part of any AI foundation.

Toward the Future of AI

We must guide artificial intelligence so AI technologies and apps conform to society's values

Context is essential for AI applications to perform accurately and responsibly, and to garner the respect and trust of users

Graphs and connections are the natural way to implement context-based AI solutions quickly and responsibly

Guidelines that promote AI should be aligned with key values of accountability, fairness and trust. Specifically, we recommend that AI systems incorporate contextual data and process for better flexibility and situational appropriateness, explainability and transparency, protection against data manipulation and verifiability.

Graph technologies are unarguably the state-of-the-art method for adding and leveraging context. These technologies are used worldwide across a diverse array of industries.

To learn more about using Neo4j for Graph Data Science, contact us today to speak to a graph technology expert.

Neo4j is the leader in graph database technology. As the world's most widely deployed graph database, we help global brands – including Comcast, NASA, UBS, and Volvo Cars – to reveal and predict how people, processes and systems are interrelated.

Using this relationships-first approach, applications built with Neo4j tackle connected data challenges such as analytics and artificial intelligence, fraud detection, real-time recommendations, and knowledge graphs. Find out more at neo4j.com.

Questions about Neo4j?

Contact us around the globe:
info@neo4j.com
neo4j.com/contact-us