KALEIDO INSIGHTS

Architecting Trust: Explainability & Ethics in AI Innovation

Fujitsu Laboratories Advanced Technology Symposium 2018
Analyst Event Summary and Point of View by Jessica Groopman

Introduction

Artificial Intelligence (AI) and machines' ability to "learn" marks a new chapter in the digital age; it breathes new life into the potential for massive and unstructured data and software, and it marks a profound shift in interface and customer experience. By any metric we are in the early days of AI, yet it is all around us. From media recommendations to navigation apps, from voice and facial recognition to cyber threat analysis, AI is powering hundreds of applications across consumer, enterprise, and government markets around the world.

Yet as we offload human capabilities onto machines, the rise of automation also calls for ethical questioning, introspection, and accountability. How do we explain outcomes and decisions? How do we prevent abuse and enable benefit to society? How do we augment human capabilities without displacement or loss of control? Most importantly, how do we foster *trust* between humans, businesses, and machines?

With growing concerns of personal data use, algorithmic bias and discrimination, AI used in warfare, and threats to individual, business, and societal health and safety, organizations must proactively assess distrust and ethics like never before.

Event Overview

In an effort to address these most imminent issues of our time, Fujitsu Laboratories Advanced Technology Symposium 2018 (FLATS 2018) focused squarely on explainable AI, the ethics of AI, and the implications and applications for business and industrial environments. The event included keynotes from leading AI researchers, Dr. Tomaso Poggio, Professor at the Dept. of Brain & Cognitive Sciences at MIT and Dr. Mark Nitzberg, Executive Director, Center for Human-Compatible Artificial Intelligence, UC Berkeley, as well as a diverse array of industry and academic leaders to share research findings, ideas, and best practices in a fascinating full-day conference attended by over 400 people in Santa Clara, California.

AI’s Rapid Resurgence Met With Numerous Impediments And Societal Questions

Although some 80 percent of organizations say they are running some form of AI in production today, companies and employees are struggling to realize widespread value beyond single point applications. Challenges resonate from within and without: Data scientists stress the need for quality (and quantity of) data; security admins race to arm ever-dynamic network topologies; IT teams scramble to update infrastructure and adapt data governance standards; executives want business results; legal teams and regulators demand auditability; employees’ navigate new tools with concerns of backlash, job displacement, and beyond.

All of this is set against the backdrop of hype about AI’s resurgence, resulting in a fragmented and chaotic landscape of vendors, academic, and government efforts, never mind inconsistent regulatory regimes, general confusion from media and society, and unprecedented ethical questions. While new technologies are always met with skepticism, AI is unique in that its very basis —understanding and reproducing human cognition—renders it subject to over-inflated expectations and human vulnerabilities.

Machine Intelligence At Scale Calls For Explainability & Ethics At Scale

The growth of AI has set into motion a realization that businesses cannot sacrifice accountability with automation. The subject of accountability in AI is vast, but boils down to two essential areas every business must address: explainability and ethics.

Explainability: The ability to see “inside” machine (and especially deep) learning networks to understand why an outcome was produced—not to mention which factors, layers, dimensions, and nodes carried the greatest weight in the decision-making— remains opaque and poorly understood. As David Gunning, program manager at the Information Innovation Office with DARPA and panelist at Fujitsu’s event put it, “we can get outcomes, but we can’t ask ‘Why that outcome? Why not another outcome?”

At the symposium we heard from representatives from FICO, PwC, Stanford University and Fujitsu on industrial requirements for explainable AI. The lack of machine introspection is problematic from the enterprise perspective in terms of low accountability, regulatory compliance, anti-discrimination, consumer protections, and erroneousness in the model. It makes models difficult to fix, tune, and de-bug. It is also problematic for external parties with an interest in knowing whether such organizations are overtly or inadvertently behaving nefariously or irresponsibly.

This “black box” challenge is a dynamic one too, as data, users, metrics, regulations, and security needs are constantly evolving. Furthermore, panelists discussed the costs (of compute, competitive exposure, and inaccuracy) of explaining every AI decision, and potential performance trade-offs companies must navigate. “Explainability is the #1 challenge to deploying AI," says Ryan Welsh, CEO of Kyndi and panelist at the Fujitsu event, citing more than a hundred interviews he'd conducted with Fortune 500 CEOs.

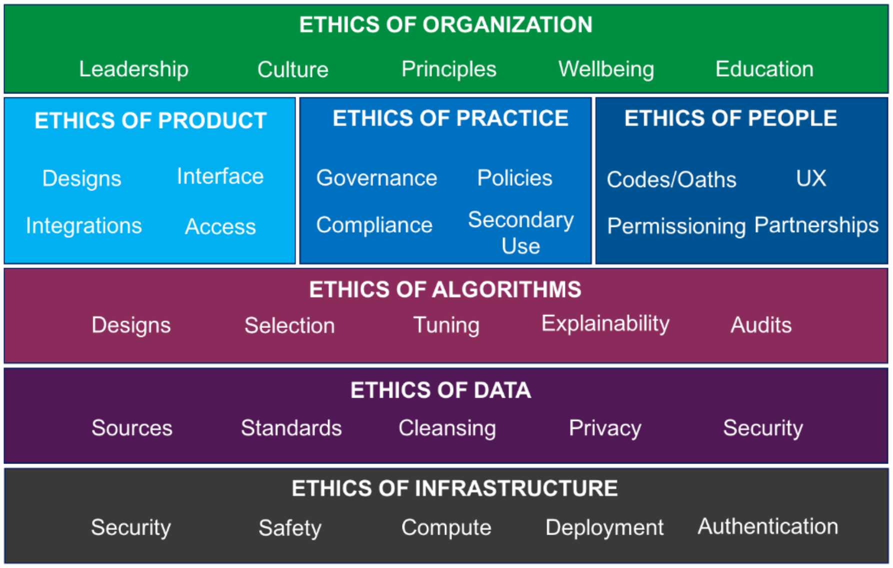
Today the need for explainability is acute in financial services, healthcare, and other highly regulated industries, but our research finds explainable AI benefits all industries because it leads to better decision making, accountability, new ways of thinking, and improved customer satisfaction. To succeed in the digital world, companies need their employees to co-create with and manage AI. "**Explainability is essential to bridge trust** between [multiple types] of user(s) and machines," says Ajay Chandler, director of Fujitsu's own Digital Life Lab which has a practice dedicated to explainability.

Ethics: Ethics are moral standards we rely on when we make decisions. As AI underlies machines' abilities to perform tasks that hitherto required a human to execute successfully, myriad ethical issues emerge. These issues tend to roll into broad societal questions around human agency, freedom, identity, access, public health, and nefarious manipulation.

From an organizational perspective, these represent diverse and difficult-to-foresee risks such as bias, discrimination, privacy, transparency, consent, compliance adherence, customer wellbeing, broken trust, and beyond. Just this year, many of the world's AI leaders have been embroiled in ethical breaches, from the Facebook-Cambridge Analytica scandal to death via self-driving Uber. The onslaught has sparked international discourse, regulatory attention, and employee backlash such as a petition signed by 4000 Google employees to never build warfare technology.

Addressing this enormous challenge means mitigation and solutions must span the proverbial business stack.

FRAMEWORK- Integrate Ethics Into the Business Technology Stack



ETHICS OF ORGANIZATION

Leadership Culture Principles Wellbeing Education

ETHICS OF PRODUCT

Designs Interface Integrations Access

ETHICS OF PRACTICE

Governance Policies Compliance Secondary Use

ETHICS OF PEOPLE

Codes/Oaths UX Permissioning Partnerships

ETHICS OF ALGORITHMS

Designs Selection Tuning Explainability Audits

ETHICS OF DATA

Sources Standards Cleansing Privacy Security

ETHICS OF INFRASTRUCTURE

Security Safety Compute Deployment Authentication

Source: Architecting Trust: Explainability & Ethics in AI Innovation Whitepaper

KALEIDO INSIGHTS

AI is too important to get wrong from a business perspective. Its importance to the future of just about every industry (thus to the bottom line) is a driving force behind these efforts. More and more, organizations that fail to design for ethical AI will be ensnared in these issues.

AI Vendors Are Approaching AI Explainability And Ethics Differently

It is no surprise that the rise of AI has simultaneously brought about a range of commercial efforts to address explainability and ethics. Within the last 12 months we have seen the largest AI technology companies in the world race to formalize all manner of ethically motivated programs, principles, and products.

Programs: Perhaps most common step companies have taken is to join industry consortia such as the Partnership on AI, interdisciplinary working groups designed to develop research, best practices, and public discourse on AI in society. Google, Amazon, Microsoft, Facebook, IBM, Apple, and others have all joined such groups, while others are building out Chief Ethics Officers or Ethics boards internally. Facebook, Microsoft, and law enforcement weapons manufacturer Axon have all assembled teams dedicated to addressing AI Ethics in the products they're building.

Principles: While most companies purport a mission or guiding values, virtually none have principles associated with AI. It was only in June of 2018, that Google, one of the world's leading organizations in the development of AI for nearly a decade, published seven principles "that actively govern our research and product development and will impact our business decisions." Since then, Microsoft, Uber, GE and others have all published their own AI principles. While principles are an essential starting point, their impact is limited without commitment, processes, and a reassessment of incentives.

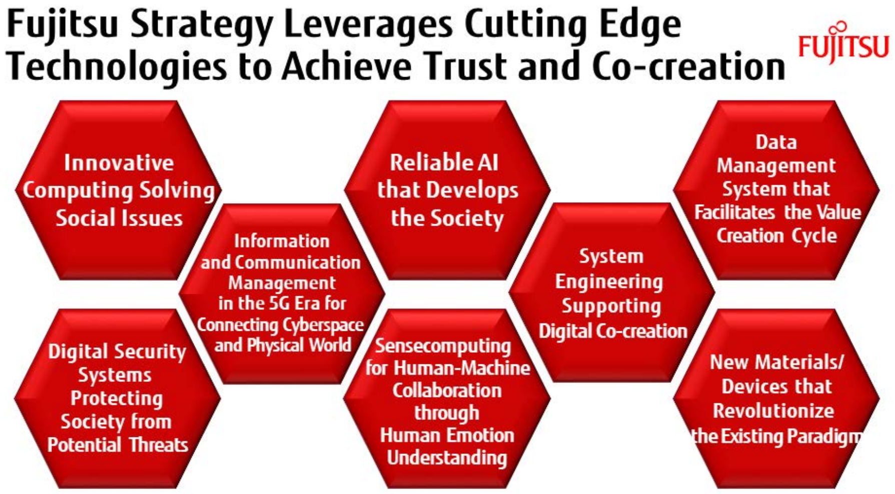
Product: Numerous companies are also working to address ethical and explainability challenges in products themselves. Accenture's Fairness Tool uses statistical methods to identify when groups of people are treated unfairly by an algorithm by analyzing how performance and "predictive parity" relate to sensitive variables. Both Facebook and Microsoft recently announced new tools designed to analyze data used to train AI systems and measuring it for particular biases for or against particular groups of people. Meanwhile IBM is supporting the communities efforts in bias recognition by releasing the world's largest annotated dataset with geo-tags and equal distribution across skin tones, genders, and ages to balance the source material and reduce sample selection bias.

Interestingly, these efforts have also inspired a growing industry of start-ups (e.g. Kyndi, Cognitive Scale, DarwinAI) and institutional efforts (e.g. LIME, Generating Visual Explanations, or DARPA's XAI) focused on tackling the many ethical and explainability gaps unaddressed by the above efforts.

Of course, abstract ethical statements, new hires, and point solutions will get us only so far. What's needed are frameworks and approaches that account for the wide range of application contexts, business models, and user types. The ethical issues involved in the use of AI for employment screening or recidivism scores, for example, are different than the ethical issues involved in the use of autonomous weapons, or even medical diagnostics. So too are the needs of the different personae in these domains to understand and trust the technology, from consumer to executive, from data scientist to doctor.

Fujitsu Anchors AI Strategy in the Architecture

What differentiates Fujitsu's strategic efforts here is that it is investing and partnering to address these issues by developing numerous innovative techniques, not just across the tech stack, but across multiple types of users and customers.


Fujitsu Strategy Leverages Cutting Edge Technologies to Achieve Trust and Co-creation
FUJITSU
Innovative Computing Solving Social Issues
Information and Communication Management in the 5G Era for Connecting Cyberspace and Physical World
Reliable AI that Develops the Society
System Engineering Supporting Digital Co-creation
Data Management System that Facilitates the Value Creation Cycle
Digital Security Systems Protecting Society from Potential Threats
Sensecomputing for Human-Machine Collaboration through Human Emotion Understanding
New Materials/ Devices that Revolutionize the Existing Paradigm

On site at the event, the company showcased fifteen examples. These examples went beyond demonstrating advanced computations; they exhibited how tech can enable trust and co-creation through knowledge-sharing, rapid decision-making, explainability, and a complementary relationship between human and AI. Two stand-outs below:

"Wide Learning" preserves transparency without sacrificing performance. A breakthrough technology, wide learning tackles two critical business hurdles. First, it compensates for insufficient training data by learning, mapping, and prioritizing hypotheses, meaning that it extracts those hypotheses worth evaluating. Second, it offers relatively greater insight and explanation into the decisioning behind outcomes. Since hypotheses are recorded as logical expressions (e.g. Women between 25-35 with an income of $50k or higher will purchase), understanding the reasoning behind a judgement doesn't require a data science degree!

"Accessible Deep Tensor" applies GUI visualization to improve accountability and accessibility. A critical extension of Fujitsu's proprietary Deep Tensor, "Accessible" Deep Tensor introduces an integrated graphical user interface (GUI) that enables engineers to filter, visualize, and configure models and systems without having to edit interconnected configuration scripts. An improved dashboard may sound incremental, but persona-defined interfaces are a critical requirement for AI explainability. Indeed, the showcase featured other examples tailored for salespeople, assurance experts, and a promising effort that connected Deep Tensor with knowledge graphs built on academic literature and genomic medicine to explain medical diagnostics in plain language.

Key benefits include:

- Explainable solutions target multiple personas (not just data scientists)
- Targeting [quantity of] training data needs helps democratize AI (not just for data-rich organizations)
- Innovative techniques benefit the ecosystem (not just Fujitsu internally)
- Outcomes prioritized for trust from R&D phase (not after-the-fact once in production)
- Improved editing and configuration tools enable stronger AI management and governance

In addition to examples outlined above, Fujitsu has also partnered with some of the world's leading research organizations working on these issues, such as MIT's Center for Brains, Minds, and Machines, INRIA, University of Oxford and Stanford University.

AI Adoption: Key Actions Towards Trustworthy AI

Although AI has been around for decades, its recent resurgence and explosive commercial and public sector application is emerging in a time of global trust erosion. Trust in government and media institutions is lower than it has been in decades, according to the 2018 Edelman Trust Barometer. Furthermore, even trust in technology companies has suffered in 2018 with revelations of cyber-breaches, election meddling, and threats to public health and safety. As a society, we are at an inflection point; now is the time when *ethics by design* will either make (or break) AI's continued commercialization.

- **Responsible AI is a business imperative.** Companies don't just require explainability for the sake of visibility; the need to interpret, question, audit, and improve models has direct financial implications. Compliance adherence, for example can cost businesses millions; erroneous outcomes resulting in harm trigger expensive litigation, nevermind brand backlash, employee attrition, and customer abandonment that take years to repair. Furthermore, explainable AI enables human explainability, a key skill for testing inferences, intuition, and unearthing new business opportunities.

- **Co-creation and collaboration are no longer an option.** As we shift from a materials-based economy to an information-based economy where intelligence is the monetizable asset, companies must turn to their ecosystems for growth. Collaboration and consensus on standards (ethical, technical, industry); on data controls systems design; on best practices and techniques are strategic. Closed, proprietary-only models are obsolete in the digital age.

- **Persona must dictate approach to AI explainability.** What constitutes an understandable explanation for *why this outcome* is dramatically different depending on *who* is interpreting *and* their *relationship to the model* itself. The explainability needs of a data scientist are distinct from a software engineer, as from an executive, a service agent, a consumer, lawyer or regulator. Needs also vary depending on what phase of the AI 'lifecycle' is being assessed: during training; editing; tweaking or tuning; learning, etc.

Dynamism, and the need to make sense of ever-increasing volume, variety, and velocity of data is the paramount business objective and principal driver of artificial intelligence. As AI evolves from *augmenting to automating* intelligent decision-making, data is indeed the fuel, but transparency and trust are the engine.

About this paper: Kaleido Insights and Fujitsu collaborated in the development of this report. Upon attending the Fujitsu Laboratories Advanced Technology Symposium 2018, Kaleido Insights analyzed the event based on its expertise in AI, digital ethics, and related categories to offer a third party review in context of the broader market. Kaleido Insights led the report development process while soliciting collaboration and insights from Fujitsu's executives, Laboratories team, and event participants. To learn more about how Kaleido Insights works with technology suppliers and adopters, or to access additional research reports, visit https://www.kaleidoinsights.com/ or reach out to julie@kaleidoinsights.com

About Jessica Groopman: Jessica Groopman is an industry analyst and founding partner at Kaleido Insights, where she leads Kaleido's automation practice and specializes in AI, blockchain, IoT and digital ethics and convergence across these areas. Jessica is a frequent speaker at emerging tech industry events and also a frequent contributor to numerous blogs and/media outlets. She has been principal analyst with Tractica, Harbor Research, and Altimeter and has served as a contributing member of the International IoT Council, the IEEE's Internet of Things Group, the DigiGuru Network, and was included in Onalytica's list of the 100 Most Influential Thought Leaders in IoT.

KALEIDO
INSIGHTS