

# Boosting AI Adoption with AI Governance

The Pillar to Safe & Responsible AI Adoption

---

Ben Eum

Advisor / AI Asia Pacific Institute

Co-founder / Armilla AI, Inc.


2021.12



**AI ASIA PACIFIC**  
INSTITUTE



**ARMILLA**



**AI will animate the world  
that the internet connected  
and electricity illuminated**



Bobby Vankavelaar



**EXCLUSIVE**

abc

**INVESTIGATION FOCUSED ON TESLA AUTOPILOT**

**abc ACTION NEWS**

11:02 83°

## The Oppau explosion

Popular Mechanics Magazine 1921



PART OF THE RUINS OF OPPAU AFTER THE DISASTROUS EXPLOSION

THE wreckage, September 21, by explosions, followed by fire, of the great dye works at Oppau, near Ludwigshafen on the Rhine, when several hundred persons were killed and thousands injured, was the greatest disaster of its kind that has ever occurred in Germany, and probably in the world. The entire plant was destroyed, as well as the greater part of the surrounding town. The first explosion occurred at the huge gas holders, and the above picture shows the resulting wreckage in their immediate vicinity. Seismographs at Stuttgart Observatory, some 25 miles away, registered the shock of the first explosion.

## Example : Governance in Chemical Engineering



ASTM INTERNATIONAL

## The cost of uncontrolled AI is high

### Ethical Breach

*Amazon's holy 'grail' recruiting tool was actually just biased against women.*

QUARTZ

### Ethical Lack of Transparency

*Lack of transparency could be AI's fatal flaw.*

△ MIND AI

### Regulatory Risks

*JPMorgan to pay \$55 million to settle mortgage discrimination complaints.*

The New York Times

### Lost ROI

*Apple Card investigated after gender discrimination complaints.*

The New York Times

*87%\* of AI projects fail", with "96% having quality and trust issues" due to lack of proper governance controls.*

VentureBeat

dimensional research

## AI governance is broken

**Industry** and regulators agree.

The time for auditability and accountability of models is no longer a nice to have.

*The industry needs to address the problem of interpretability [....] before putting investor money at play.*

Head of Liquidity Research

**BlackRock**

*There is no time frame for when an AI system could be deployed [....] because solving the explainability problem is so important.*

Managing Director, Bank of America

**BANK OF AMERICA**

*We only roll out machine learning when we feel comfortable there are no biases or lack of transparency.*

Managing VP, ML & AI

**CapitalOne**

*We see a rapid increase in explainability from clients and a real roadblock to moving models to production.*

Industry Principle, ML & AI

**nVIDIA**



## AI governance is broken

Industry and **regulators** agree.

The time for auditability and accountability of models is no longer a nice to have.

*Ultimately, explainability is a prerequisite to secure the very principle of responsibility.*



*Explainability may be important conditions not only for the risk management [...] but also for greater trust from the general public as well as regulators and supervisors.*



*Transparency is one of our key principles for the use of AI and data analytics [...] in the provision of financial products and services.*



*[...] We may need, as a regulator, to look under the hood or behind the curtain to see what data were used, what factors were programmed into the system and what question the AI was trained to answer.*





# Abundance in AI guidelines

Values to be respected	AI Utilization Guidelines	Draft AI R&D guidelines for international discussions	Social Principles of Human-centric AI	Ethics Guideline for Trustworthy AI	Recommendation of the Council on Artificial Intelligence	Ethically Aligned Design	Asioma AI Principles	Tenets
by	The Conference toward AI Network Society (MNC) /Japan	The Conference toward AI Network Society (MNC) /Japan	Integrated Innovation Strategy Promotion Council, (Council after Social Principles of Human-centric AI) /Japan	European Commission (High Level Expert Group on AI (AI HLEG))	OECD	IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	Future of Life Institute (FLI)	Partnership on AI
Released on / Previously released on (Draft doc. ver.)	2019/9/9 2019/7/17 (Draft AI Utilization Principles)	2017/7/28	2019/3/29 2018/12/27 (Draft)	2019/3/29 2018/12/18 (Draft)	2019/9/22	2019/3/23 (1st edition) 2016/12/13 (ver.1), 2017/12/12 (ver.2)	2017/7/	2016/9/28
URL	<a href="http://www.mnc.go.jp/mnc/_content/00057894.pdf">http://www.mnc.go.jp/mnc/_content/00057894.pdf</a> <a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>	<a href="http://www.mnc.go.jp/mnc/_content/00057897.pdf">http://www.mnc.go.jp/mnc/_content/00057897.pdf</a>
Structure	Purpose Basic philosophies (7) AI Utilization principles (10) and their comments	Purpose Basic philosophies (5) AI R&D principles (9) and its comments	2. Basic philosophies (3) 3. Vision (5) 4.2. AI R&D and Utilization Principles etc.	1. Foundations of Trustworthy AI/4 Principles(P) 2. Realizing Trustworthy AI: Requirements(R) 7.1-Technical and non-technical methods 3. Assessing Trustworthy AI, ...	Common understanding of terms 1.Principles for responsible stewardship of trustworthy AI (5) 2.National policies and international co-operation for trustworthy AI (5)	pillars (3) General Principles (GP: 8) Chapter (11 including GPs)	Principles (23)	Tenets (10)
Human-centered	<b>Basic Philosophies</b> To achieve a human-centered society where all human beings across the board enjoy the benefits from their life in harmony with AI networks, while human dignity and individual autonomy are respected.	<b>Basic Philosophies</b> To achieve a human-centered society where all human beings across the board enjoy the benefits from their life in harmony with AI networks, while human dignity and individual autonomy are respected.	<b>2(1). Dignity: A society that has respect for human dignity</b> We need to construct a society where human dignity is respected and, by using AI as a tool, a society where people can better demonstrate their various human abilities, show greater creativity, engage in challenging work, and live richer lives both physically and mentally.	<b>P1: Respect for human autonomy</b> The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process..... The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.	<b>1.2. Human-centred values and fairness</b> Governments should call on AI actors to be founded on directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process..... The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.	-	-	-
Human dignity	<b>7) Principle of Human Dignity and Individual Autonomy</b> Users should respect human dignity and autonomy in the utilization of AI systems or AI services. - Attention to the manipulation of human decision-making, emotions, etc. by AI - Reference to the discussion of bioethics, etc. in the case of linking AI systems with the human brain and body	<b>7) Principle of Ethics</b> Developers should respect human dignity and individual autonomy in the R&D of AI systems. - Encouraged that developers pay particularly due consideration to respecting human dignity and individual autonomy, in light of discussions on bioethics, etc. - Advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity. - Advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity.	<b>4.1(1) The Human-Centric Principle</b> The utilization of AI must not infringe upon the fundamental human rights guaranteed by the Constitution and international standards. AI should be developed, utilized, and implemented in society to expand the abilities of people and allow diverse people to pursue their own well-being.	<b>P1: Respect for human autonomy</b> The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process..... The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.	<b>1.2. Human-centred values and fairness</b> Governments should call on AI actors to be founded on directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process..... The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.	<b>GP1. Human Rights:</b> AI/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.  <b>GP2. Well-being</b> AI/IS creators shall adopt increased human well-being as a primary success criterion for development.	<b>10) Value Alignment:</b> Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.  <b>11) Human Values:</b> AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.	<b>3) We are committed to open research and dialogue on the ethical, social, economic, and legal implications of AI.</b>  <b>4c) Maximize the benefits and address the potential challenges of AI technologies, by ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.</b>
Diversity and Inclusiveness	<b>Basic Philosophies</b> To respect users' diversity, as well as an inclusiveness of people with diverse backgrounds, values, and ideas, throughout the process of the utilization of AI.	<b>Basic Philosophies</b> To achieve a human-centered society where all human beings across the board enjoy the benefits from their life in harmony with AI networks, while human dignity and individual autonomy are respected.	<b>2(2) Diversity &amp; Inclusion: A society where people with diverse backgrounds can pursue their own well-being</b> It is both an ideal in the modern world and a major challenge to create a society in which people with diverse backgrounds, values and ways of thinking can pursue their own well-being while society realizes new value by flexibly embracing them.  <b>4.1(1) The Human-Centric Principle</b> In the process of AI deployment, each stakeholder should take into consideration the user-friendliness of the system in order to allow all people to enjoy the benefits of AI and avoid creating a digital divide (the so-called "information poor" or "technology poor" people left behind).  <b>4.1. (2) The Principle of Education/Literacy</b> In a society premised on AI, we do not desire to create disparities or divisions between people or create those who are socially disadvantaged.	<b>RS: Diversity, non-discrimination and fairness</b> In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness.  <b>4.1. (2) The Principle of Education/Literacy</b> In a society premised on AI, we do not desire to create disparities or divisions between people or create those who are socially disadvantaged.	<b>1.1. Inclusive and sustainable growth and well-being</b> Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as empowering human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.	<b>14) Shared Benefit:</b> AI technologies should benefit and empower as many people as possible.  <b>15) Shared Prosperity:</b> The economic prosperity created by AI should be shared broadly, to benefit all of humanity.  <b>23) Common Good:</b> Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.	-	-

## | Two practical initiatives we can do now

At individual **enterprise** level...



**Incorporating  
Model Validation  
into ML Workflow**

At **industry** level...



**'Failing Forward'  
to Collect & Share  
Learnings**

## Automated model validation is no longer a nice to have

### **Business decisions are increasingly model-driven**

The increasing volume and impact of decision-support models is becoming the primary source of both company value and risk.

### **Model size and complexity is growing fast**

Increasing complexity, sophistication, and diversity of models, modelling techniques, and platforms creates brand new challenges for testing and QA teams.

### **Model testing and QA processes are not keeping pace**

Most standard testing and QA practices are unable to keep up with model development or to effectively balance rigor with speed.

## Industry is paying high price

**6+**  
**months**

To validate and  
release a model

**87%**  
**Of AI projects fail**

1 in 3 models are  
abandoned or shelved due to  
lack of transparency

**\$7+**  
**billion\***

Of estimated value  
currently locked due to  
gaps in AI Governance

**Poor model  
accountability**

Results in faulty AI that  
makes biased and  
unethical decisions

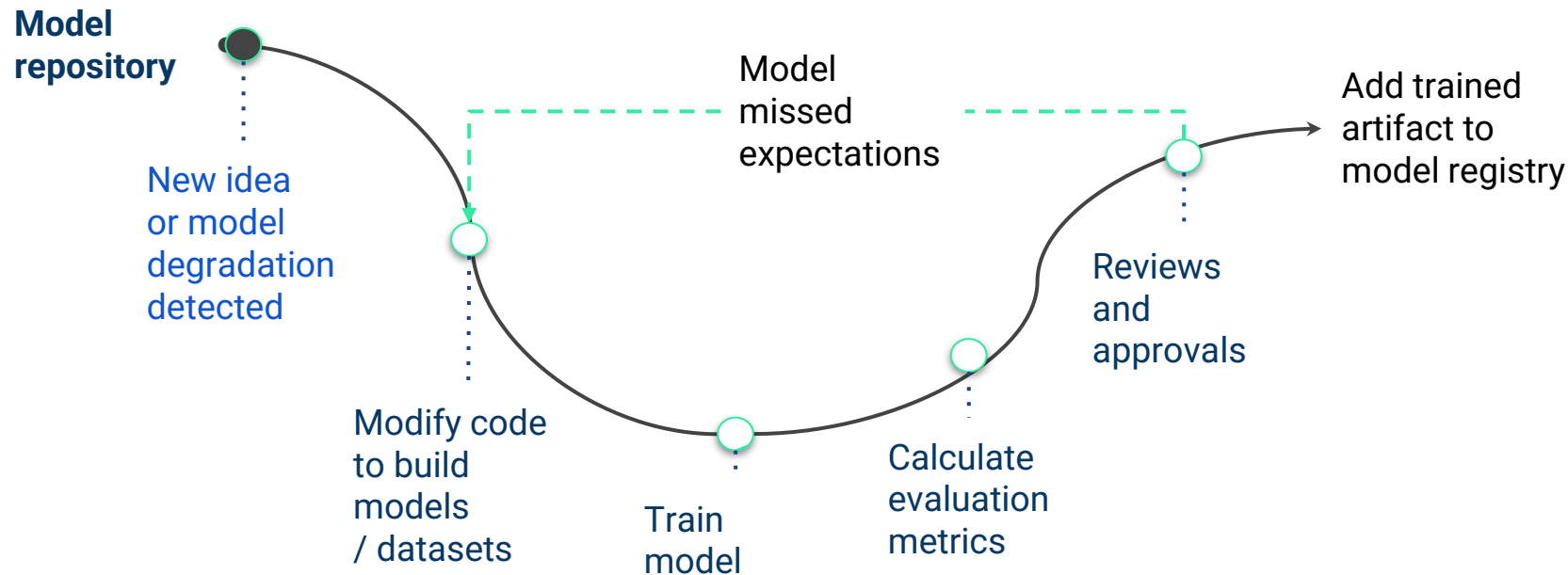
### Current ML testing

- Siloed
- Inefficient
- Risky

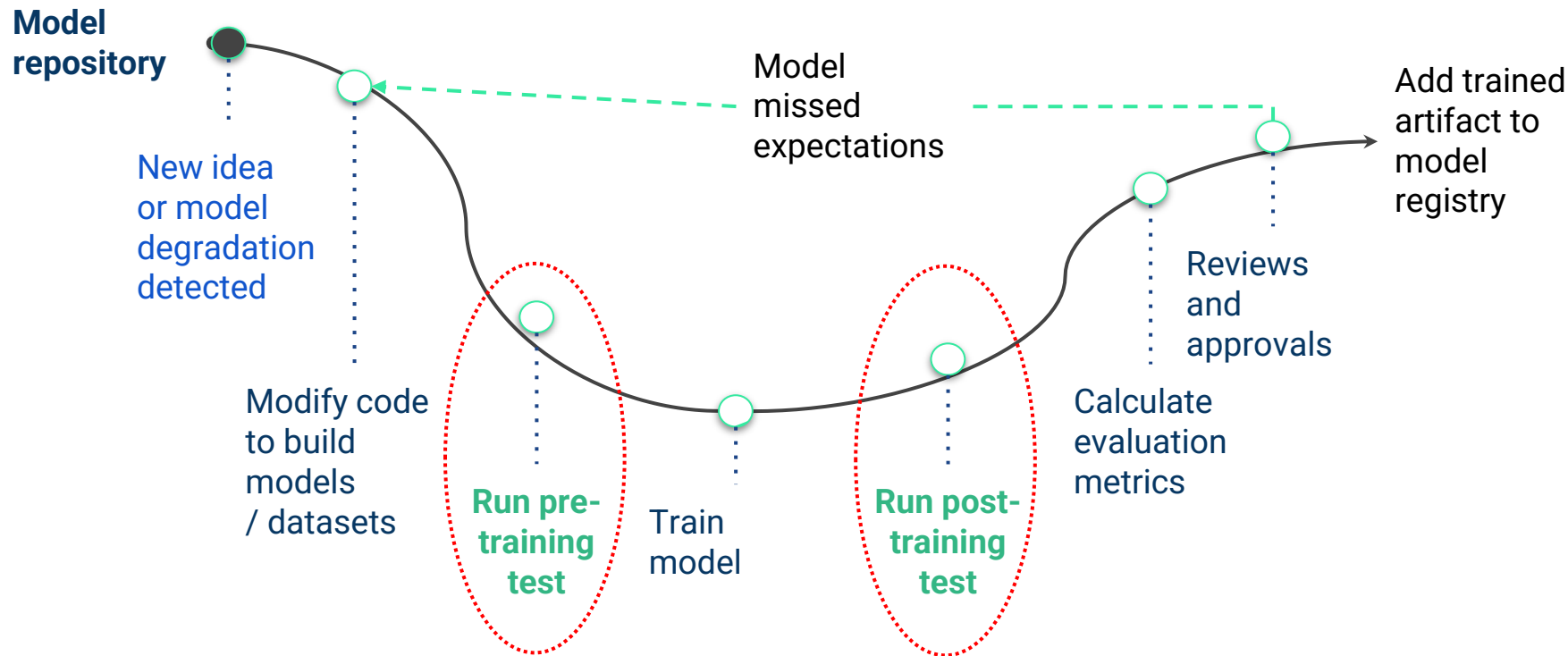
The integration of AI requires  
a new approach to model  
testing and risk management.



## Typical workflow of (current) model development – notice the lack of ‘validation’



## A workflow for developing high-quality models



## Validation must be 'STRATEGIC'



**The TOURING**  
\$290  
F. O. B. Detroit

Color—Black. All-steel body with pronounced streamline effect. Low, deeply cushioned seats. Double ventilating windshield. Our sun top. Standard equipment includes weatherproof floor, curtains opening with all four doors, windshield wiper, nickel-plated headlamps, rimmed four cord tires. Starter and demountable rims, \$85 extra. Radio, \$25 extra.

**The FORDOR SEDAN**  
\$660  
F. O. B. Detroit

Color—Wardens Victory to harness. Nickel-plated chrome. Double ventilating glass windows and all equipment in rear windows, do view. Windshield wiper, dash lamp rim and four cord tires.

**The RUNABOUT**  
\$260  
F. O. B. Detroit

Color—Black. All-steel body with streamline treatment. Large compartment under the steering, rear deck. Double ventilating windshield. Standard equipment includes weather-proof floor, curtains opening with both doors, windshield wiper, nickel-plated headlamps, rim and four cord tires. Starter and demountable rims, \$85 extra. Radio, \$25 additional.

**The COUPE**  
\$520  
F. O. B. Detroit

Color—Charcoal Green with spoked body. Nickel-plated radiator and hood. Windshield and hooded sun visor rotary lift. Wide shell behind seat. Rear deck compartment for luggage. Radio, \$25 extra.

**Ford**



## What is 'Failing Forward'

'Pay It Forward' (2000, Mimi Leder)





**Purpose : sharing 'experiences' to boost 'collective learning & consensus'**

**Fail Fast**



**Learn Faster**

**Fail Forward**



**Learn Together**

## Inspirations



## Call for participation : AI Asia Pacific Institute



### Pan-Asian research on sharing failures in AI governance

As more organisations adopt AI technologies, 'governing' AI to build trust between human and machine is critical. One way to promote trustworthy AI is to make past failures known, so that AI systems can be built on trust.

The goal of this initiative is to research on the proper mechanisms and platforms for sharing relevant failures – big and small – in AI development and deployment, promoting collaboration of governments and commercial entities across Asia.





# Thank you

---

For more information, please contact:

[ben@armilla.ai](mailto:ben@armilla.ai)

+82.10.2051.8371

[armilla.ai](http://armilla.ai)



AI ASIA PACIFIC  
INSTITUTE



ARMILLA