# Building a Robust AI Infrastructure: Essential Components and Considerations

Yingquan Li

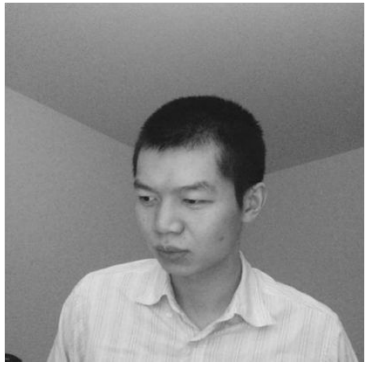
Data Engineer, Space Telescope Science Institute (STScI)

The AI Summit of New York (NY), 2024

# About Me

- Education:

- Work:



My background is:

- *Science/Engineering* and *Business*

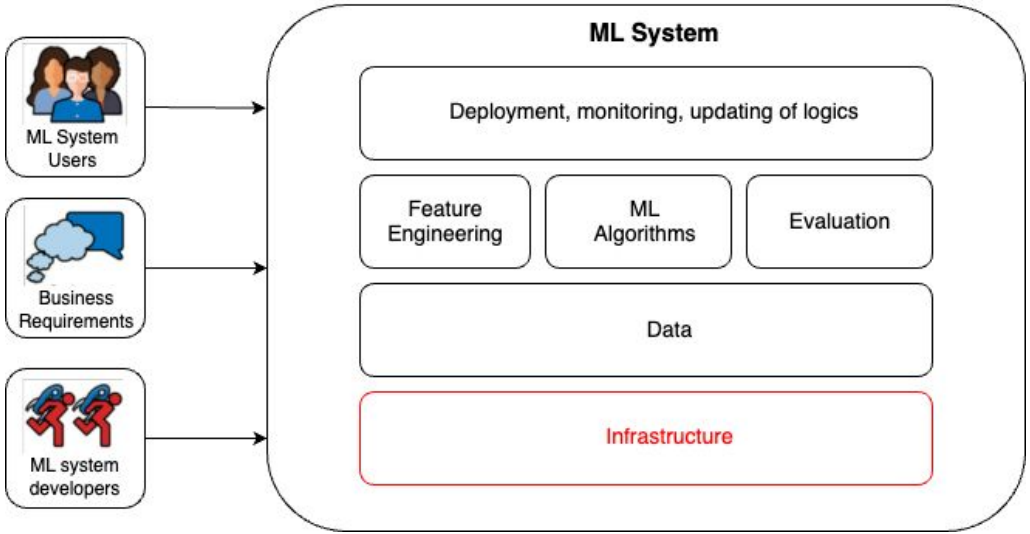
I've worked in:

- *Government*, *Academia*, *Private Industry*

# Agenda

- What is AI Infrastructure?
- AI Infrastructure: Customization and Standardization
- AI Infrastructure: Best Practices
- The Role of Data Governance and Ethics
- Conclusion

# What is AI Infrastructure?




**Chip Huyen** (*Designing Machine Learning Systems*, 2012)

- **Infrastructure** comprises *hardware*, *software*, *networking* tools that sit at the bottom-most layer of any AI/ML system.
- The ultimate goal of *any* AI infrastructure is to turn business requirements into production AI use cases.
- Key Point: Infrastructure means different things to different people (*Engineer*, *TPM*, *Principal Engineers/SME*, *VP*/*Director*/*CIO*). Remember this point when it comes to communication!

# AI Infrastructure: Customization and Standardization

**Four** key areas to consider!

# AI Infrastructure: Customization and Standardization

## #1: Infrastructure Control

Should your infrastructure be hosted ***on-premise***, on a boxed-in ***public cloud platform*** or outsourced to a ***third party (i.e. SaaS)***? What are the tradeoffs of each?

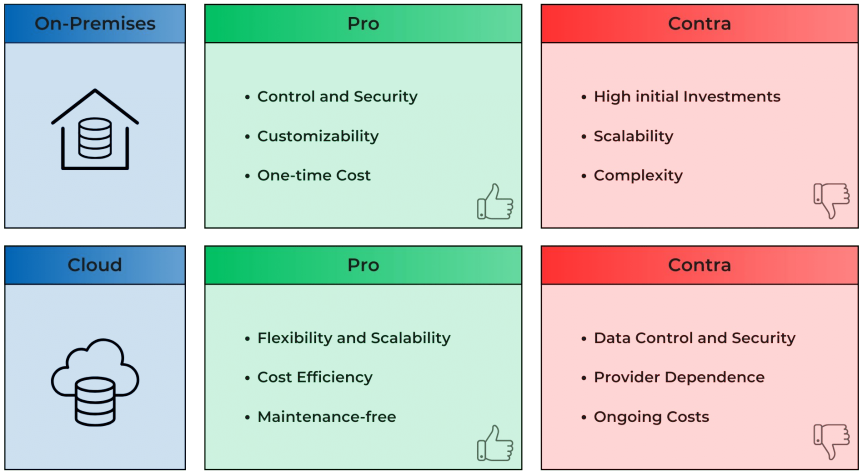



Image Source

# AI Infrastructure: Customization and Standardization

## #2: Industry Regulations

Does your company work in an industry where there are ***regulatory laws*** that are in play, such as *HIPPA - Healthcare* or *AML - Finance*?


Image Source

# AI Infrastructure: Customization and Standardization

## #3: Infrastructure Effectiveness

Is your solution ***scalable***, ***performant***, and ***secure***? Specifically, is the infrastructure itself secure and is the data that you have secure?

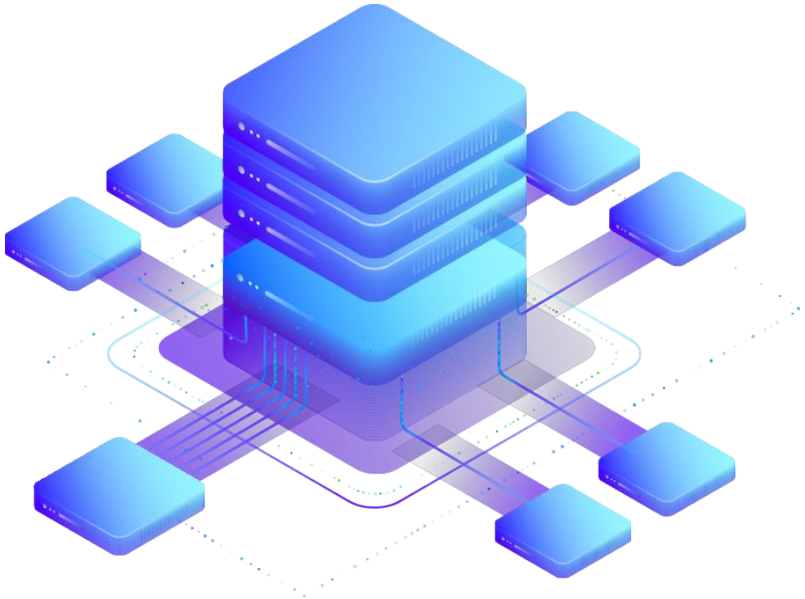


Image Source

# AI Infrastructure: Customization and Standardization

#4: **Technical + Business Objectives Alignment**

Does the infrastructure and AI being developed i.e. *technical objectives* ultimately tied back to *business objectives*? ***YES*** or ***NO***?

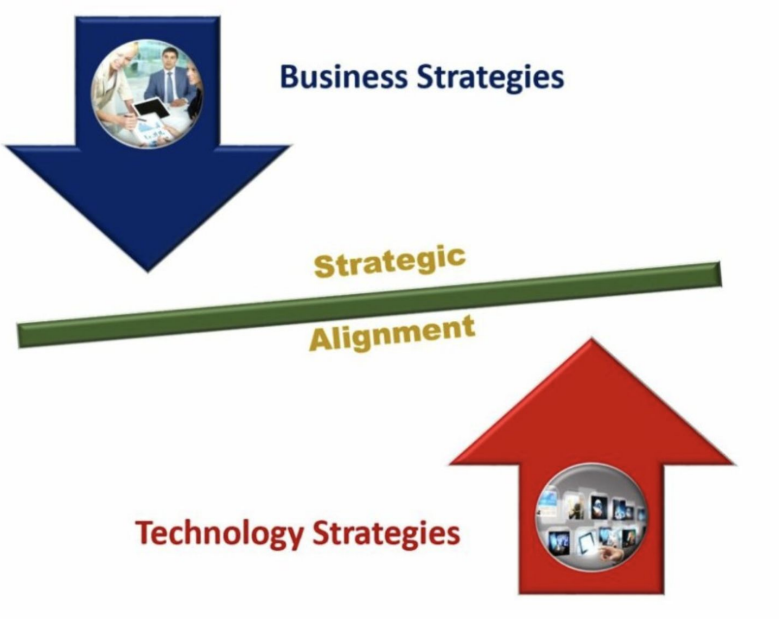



Image Source

# AI Infrastructure: Best Practices

These are the *best practices* for each of the **four** key areas!

# AI Infrastructure: Best Practices

Best Practice For: **Infrastructure Control**

Infrastructure hosted on a ***public cloud platform*** offers the best mix between customization and standardization, unless you work in the following industries: finance, insurance, healthcare or government. In these industries, infrastructure is going to be hosted ***on-premise*** and change will be INCREMENTAL.

# AI Infrastructure: Best Practices

Best Practice For: **Industry Regulations**

Please follow the rules for your industry! **Data governance** and *compliance*, something that we will talk about more later, should not be something just to check the box! As much as regulations may be a pain, they do serve a purpose and we have to get things right with the rules.

# AI Infrastructure: Best Practices

Best Practice For: **Infrastructure Effectiveness**

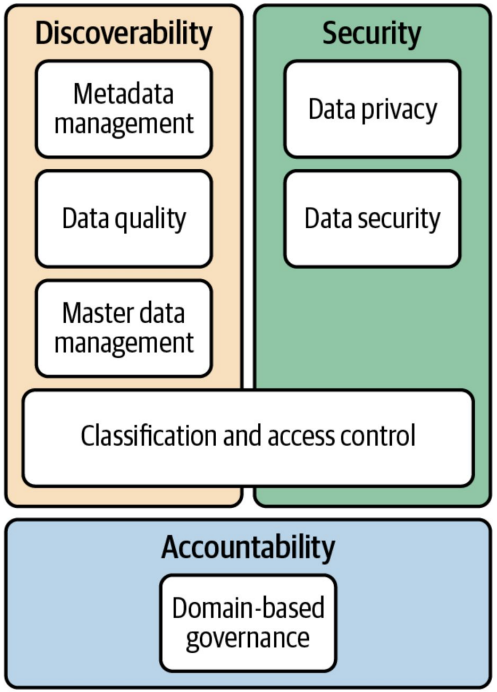
- Select technologies that have *widespread adoption*, *good support*, and is *mature*. Work with vendors who are trusted by others. If allowed, use **open-source tools** to *reduce cost*!
- *Security*, *security*, *security*! Make sure data is safely secured and PII is NEVER lost. Implement STRONG protocols around access controls.

# AI Infrastructure: Best Practices

Best Practice For: **Technical + Business Objectives Alignment**

*Communication*, *communication*, *communication*! Make sure that your engineers understand that the technical work they do **MUST** ultimately tie back to ***business objectives***. Make sure technical folks understand this by any means necessary. It's not about the technology, but the business impact!

# The Role of Data Governance




**Evren Eryurek** et al. (*Data Governance: The Definitive Guide*, 2021)

- **Data Governance** is all about building *trust* in the data.
- The three fundamental pillars for the business leader to promote: Discoverability, Security, and Accountability.
- The two fundamental pillars for the engineer to build: Classification and Access Controls.
- Key Point: Data Governance initiatives must be **top-down**, with leadership buy-in at the start! Build a real **data culture** by treating data governance *SERIOUSLY*.

# The Role of Ethics

Pres. Biden's **AI Executive Order** (Oct. 2023)

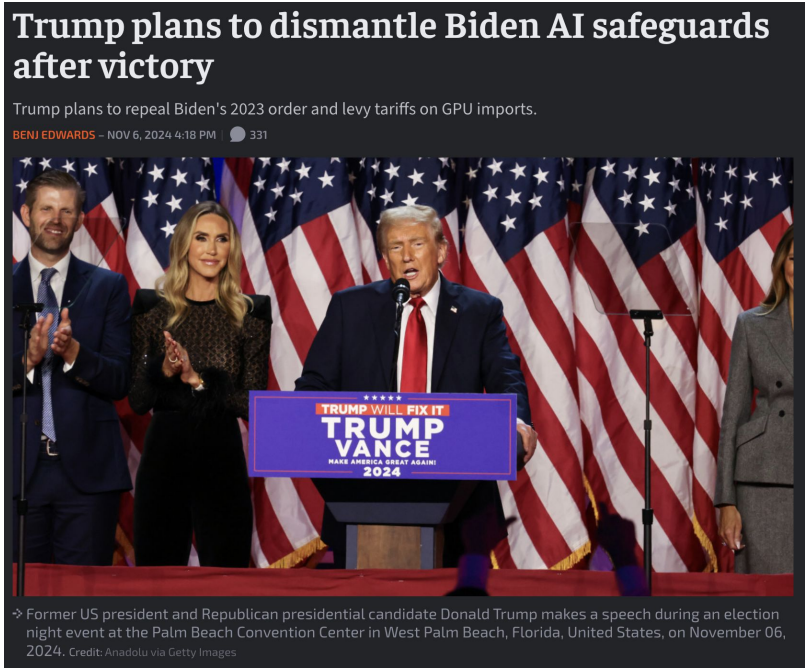
- AI developers must disclose safety test results and key information to the U.S. government.
- Create standards, tools, and tests to ensure AI systems are safe, secure, and trustworthy.
- Guard against AI misuse in creating hazardous biological materials.
- Establish standards to detect AI-generated content and authenticate official content, protecting against AI-enabled fraud and deception.
- Develop an advanced cybersecurity program to create AI tools that identify and fix critical software vulnerabilities.
- Create a National Security Memorandum for further AI and security actions.

Image Source

# The Role of Ethics

## The **Republican** Perspective

"Representative Nancy Mace (R-S.C.) warned that reporting requirements could *discourage innovation* and prevent developments like **ChatGPT**."




(*Article on **Ars Technica***)

# In Conclusion, We've Learned the Following:

**Four Best Practices**

- Infrastructure Control: Use a *public cloud platform* if you can!
- Industry Regulations: *Follow the rules*, do the right thing!
- Infrastructure Effectiveness: Use the *best technology* and keep the *infrastructure + data* secure!
- Technical + Business Objectives Alignment: Tie *technical objectives <-> business objectives*, PERIOD!

**Data Governance** and **Ethics**

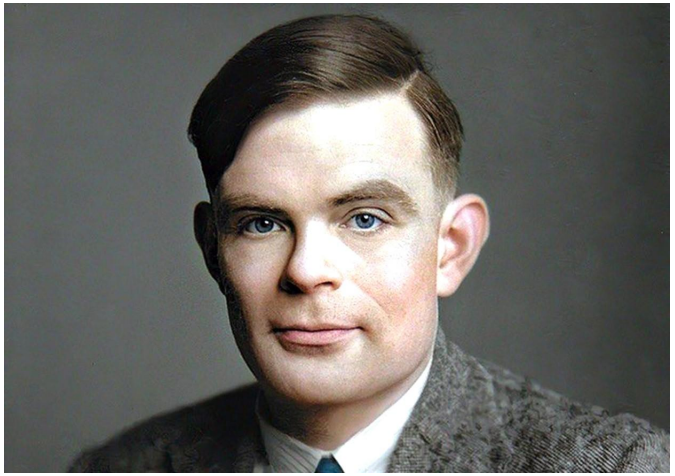
- Data Governance initiatives must come from the *top-down*, ALWAYS!

- When it comes to Ethics, how do we balance *good morals* and *innovation*? We need BOTH! The decision rests with you 🤓.

# When the **Legend** Speaks, We Listen 👑

"**We may hope that machines will eventually compete with men in all purely intellectual fields**. **But which are the best ones to start with**? **Even this is a difficult decision**. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.

**We can only see a short distance ahead, but we can see plenty there that needs to be done**."



- **Alan Turing**
*Computing Machine and Intelligence* (1950)

Image Source

# Thank you for Your Attention!

yli12313@vt.edu

301-204-3957

# Citations


- Books:
  - Huyen, Chip. (2022). *Designing Machine Learning Systems*. O'Reilly Media, Inc.
  - Evren Eryurek et al. (2021). *Data Governance*: *The Definitive Guide*. O'Reilly Media, Inc.
- Articles:
  - AI Infrastructure Solutions: SaaS vs. Self-Managed Setups (On-Prem & VPCs)
  - Trump plans to dismantle Biden AI safeguards after victory
  - FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence
- Papers:
  - Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, 59,* 433–460. https://doi.org/10.1093/mind/LIX.236.433