Accelerating AI Lifecycle

Meta's Fast ML Prototyping Insights

Parul Gupta
Senior Production Engineer, Meta

Who am I?

Parul Gupta, Senior Production Engineer at Meta.

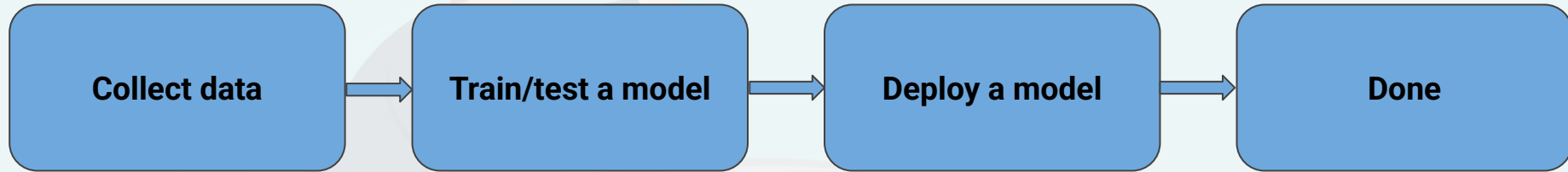Work on Python Language Foundation Team. Prev, Interactive Computing Team.
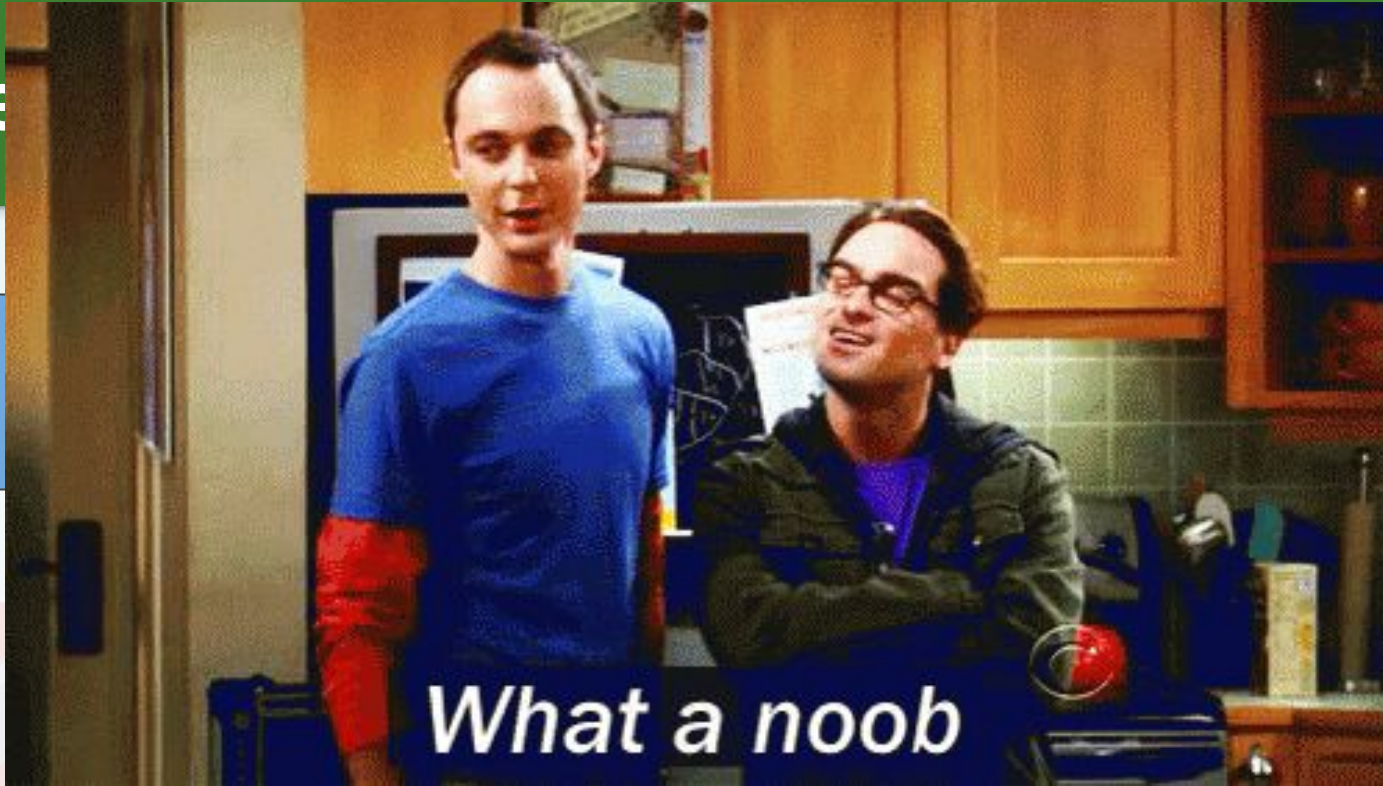
Outside, I dance and I paint.

Accelerating AI Lifecycle

AI life-cycle

Expectation

| Collect data | | Train/test a model | | Deploy a model | | Done |

Expe

Collect

Done

AI Lifecycle

Reality

Figure 2-2. The process of developing an ML system looks more like a cycle with a lot of back and forth between steps
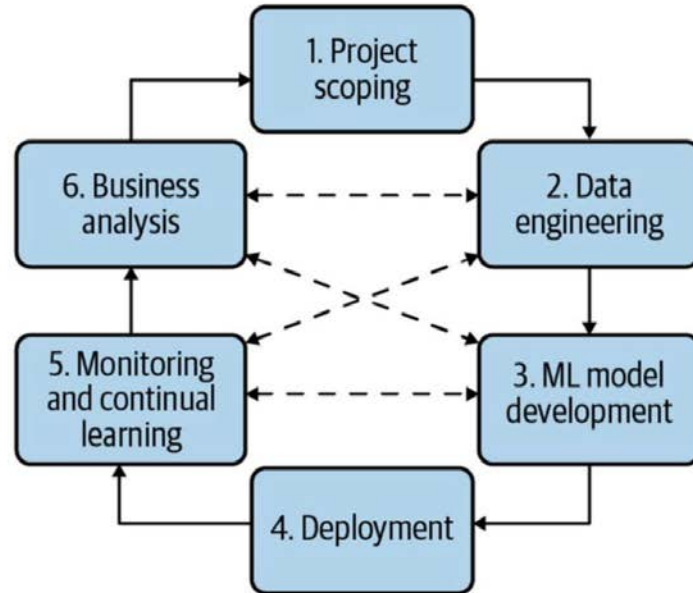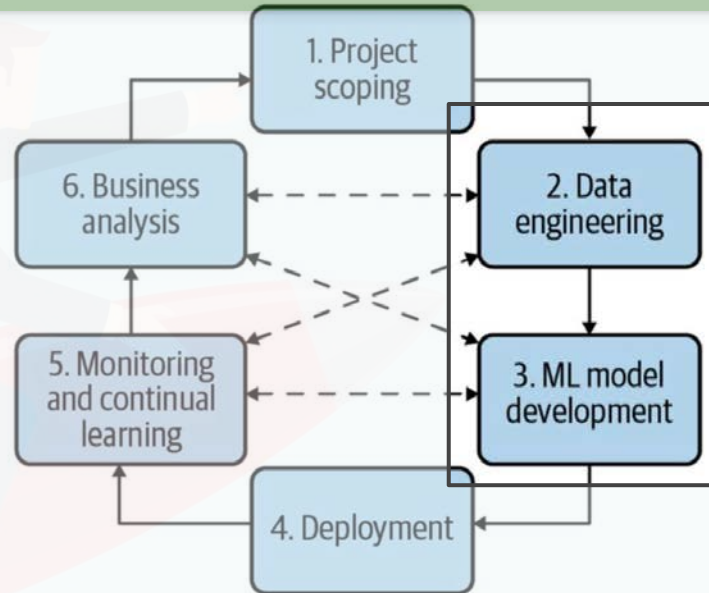
Reality

Figure 2-2. The process of developing an ML system looks more like a cycle with a lot of back and forth between steps

*(Source: **Designing Machine Learning Systems, O'Reilly Media, Inc.**)*

AI Lifecycle

ML Prototyping Tools

Jupyter notebooks

- Open-source, web-based interactive computing platform.
- Contains live code, equations, visualizations, narrative text and images
- Benefits?
  - Data science/engineering/processing
  - ML modeling and testing
  - Language agnostic
  - Backend agnostic

ML Prototyping Tools

Conda

- **Popular package manager**
- **Developed by Anaconda Inc**
- **Isolated Virtual environment/package manager**
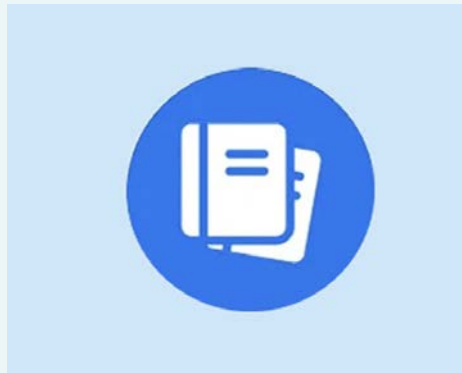

THANKS ANA-CONDA!

How we accelerated
Let's share ideas

Bento - Tech Accelerations at Meta

Bento is internal-only Jupyter-based distribution.

Optimized to support our needs.

How we accelerated? Let's share ideas

Bento - Tech Accelerations at Meta

Scheduled notebooks

https://engineering.fb.com/2023/08/29/security/scheduling-jupyter-notebooks-meta/

Bento - Tech Accelerations at Meta

Persist sessions / notebook recovery across servers

https://cfp.jupytercon.com/2023/talk/7X3G8H/

How we accelerated? Let's share ideas

Bento - Tech Accelerations at Meta

Cross platform debugging (VSCode@Meta <-> Bento)

Notebooks in VSCode@Meta

ALIGN ALL THE PLATFORMS!

How we accelerated? Let's share ideas

Bento - Tech Accelerations at Meta

Integrations to enable data privacy and security reviews

Support for serverless servers

Enhanced SQL cell support

Ability to use multi-language kernel

Auto-suggestions using internally-developed GenAI agents (e.g, CodeCompose)

And much more...

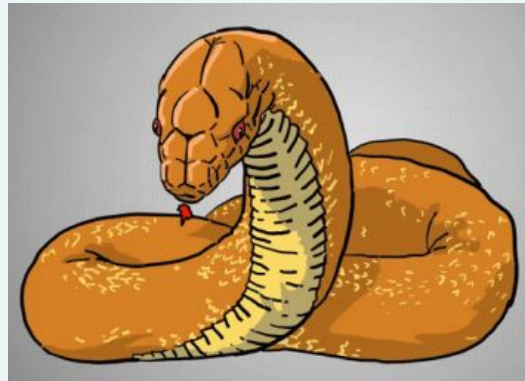*How we accelerated? Let's share ideas*

Conda - Tech Accelerations at Meta

Open source - Lower learning curve

Full version control of dependencies in our monorepo

Portable environments for training jobs

Trackable workflows

How we accelerated? Let's share ideas

Combining the two further accelerates the AI development

Thank You

Parul Gupta

https://www.linkedin.com/in/parulgupta04/