

LLMs for Agile Innovation

Josh Seltzer | Practical LLMs Workshop

conversational AI for

**deeper insight,
faster**

Agenda: Idea → Product

- Solving a problem
- Springboard
- Exploratory R&D
- LLM Microservices
- Optimization and Production Engineering
- Resources
- IP Strategy

Solving a Problem

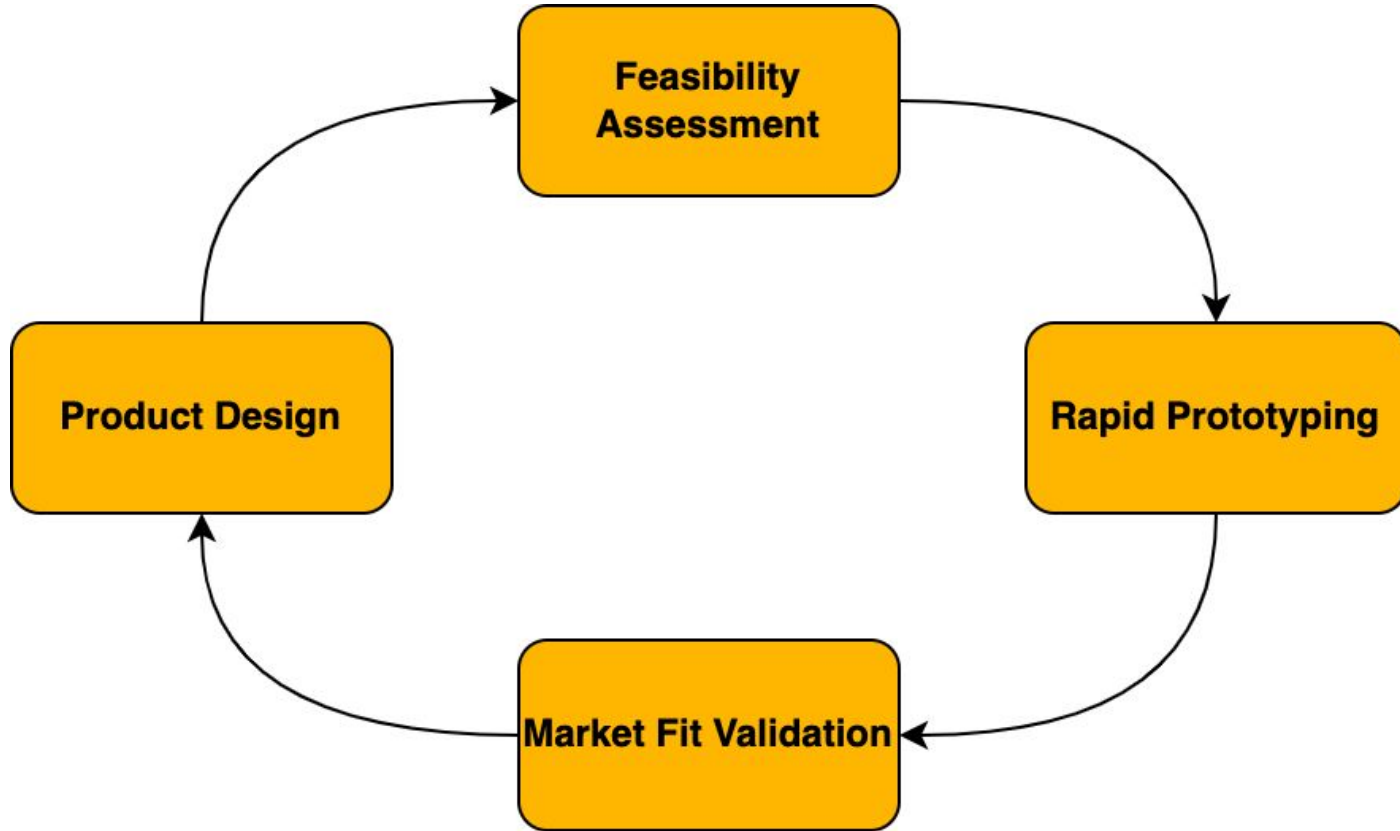
- Deeply understanding your target users is essential...
 - ... but think beyond what is conventionally possible
- LLMs, fine-tuned and applied for your domain, offer qualitatively new ways of solving problems
- Establishing quick feedback from a real user

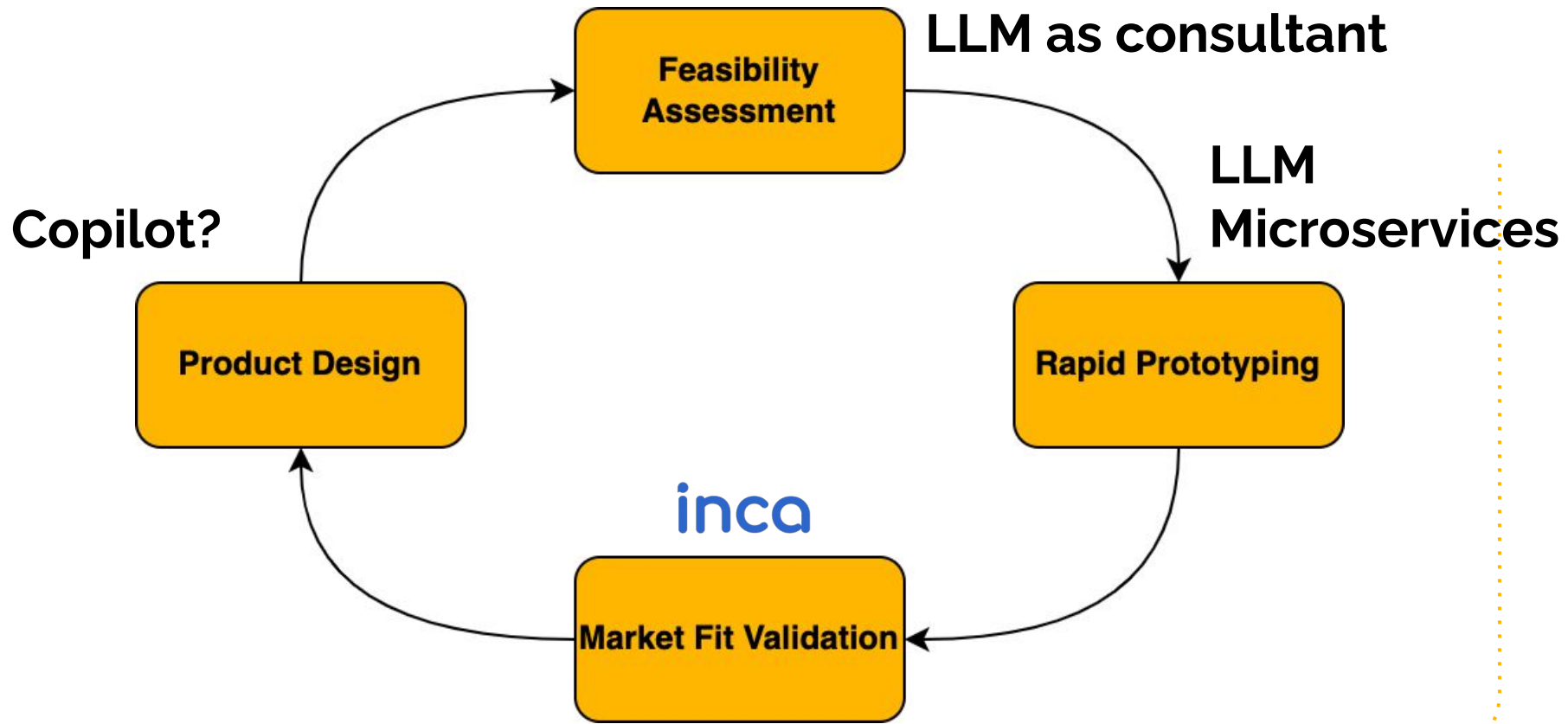
Springboard

- Hackathons and Accelerators
- Investment vs Bootstrapping

Exploratory R&D

- Focus on solving your problem through rapid iteration
- DON'T focus on cost!

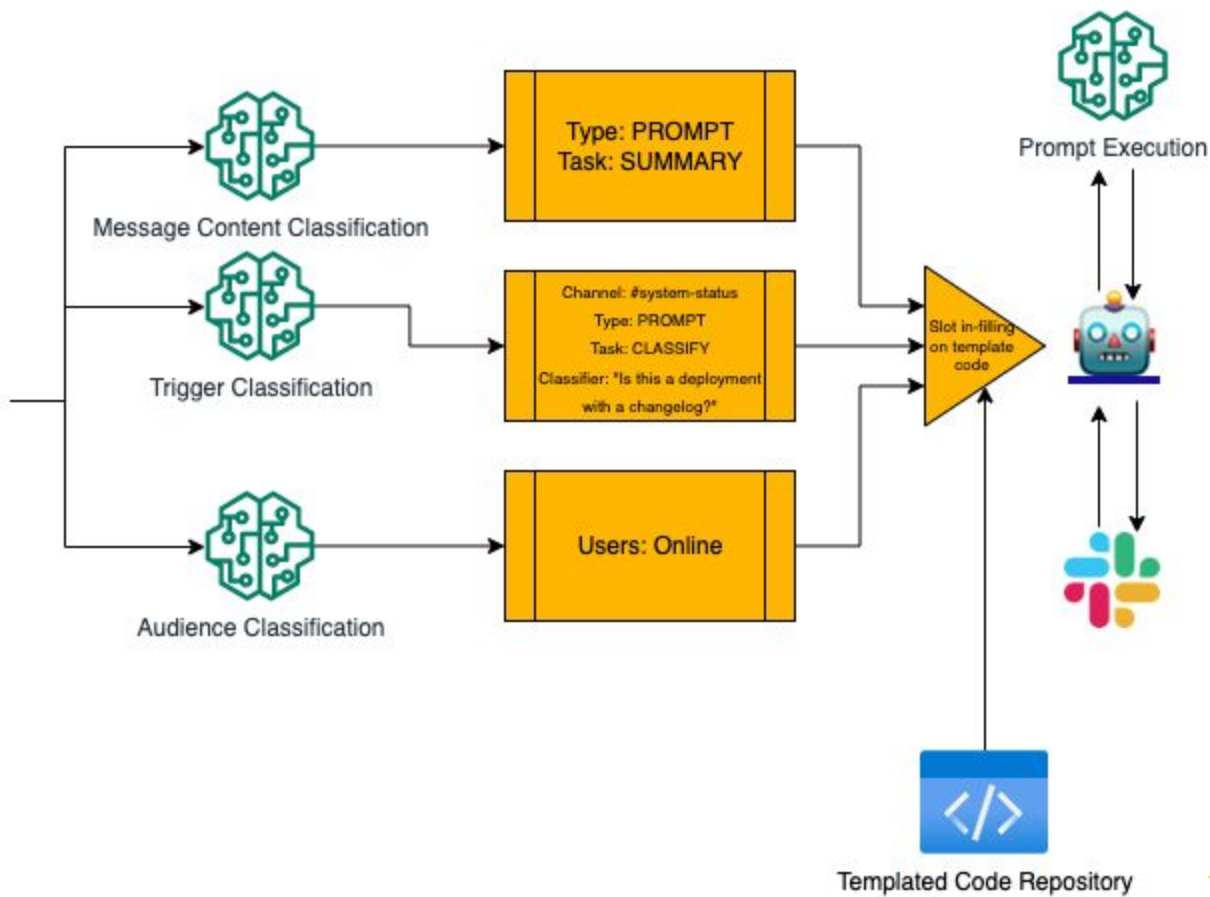




LLM Microservices

- Conquer and divide!
- LLMs for rapid prototyping
- Easier explainability, unit testing, quality control, etc.

I want a slack bot which msgs everyone online whenever there's a new deployment announced on #system-status with a summary of the release changelog.



Optimization / Engineering

- Cost and performance optimization
- Data protection and governance
- Scalability
- Fallback and quality control

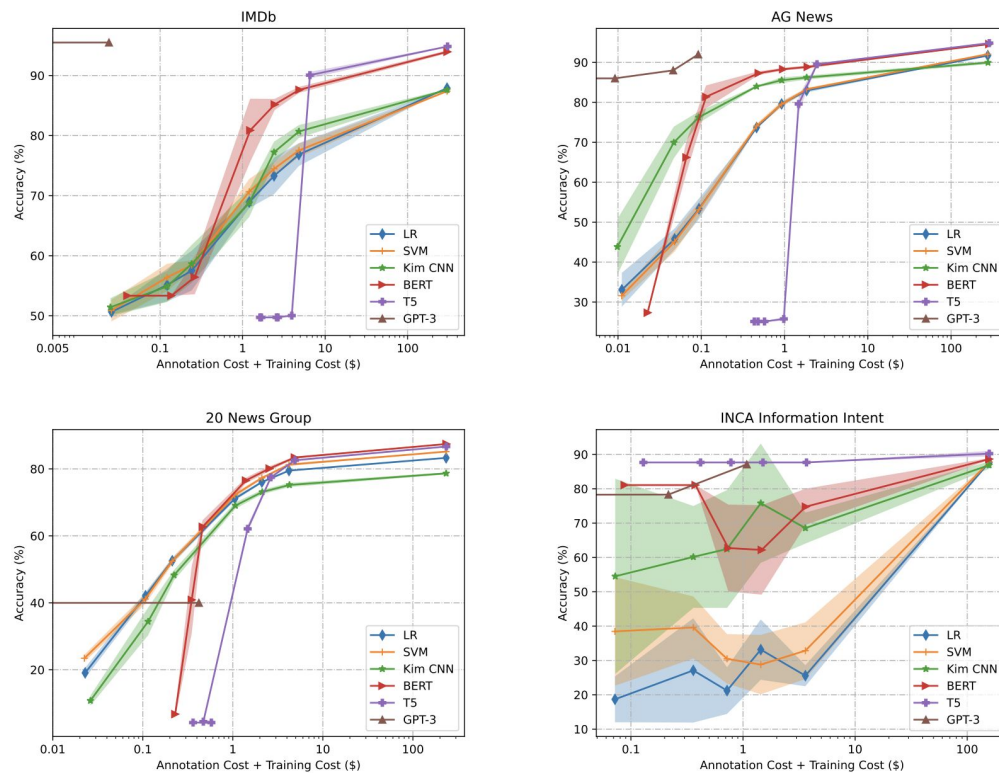
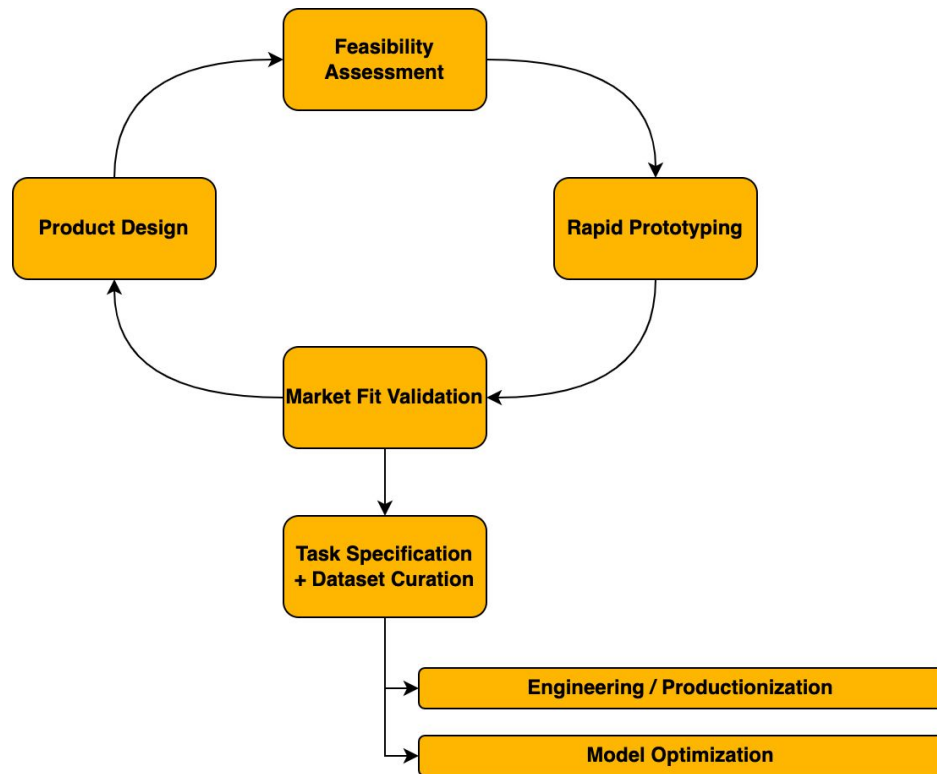
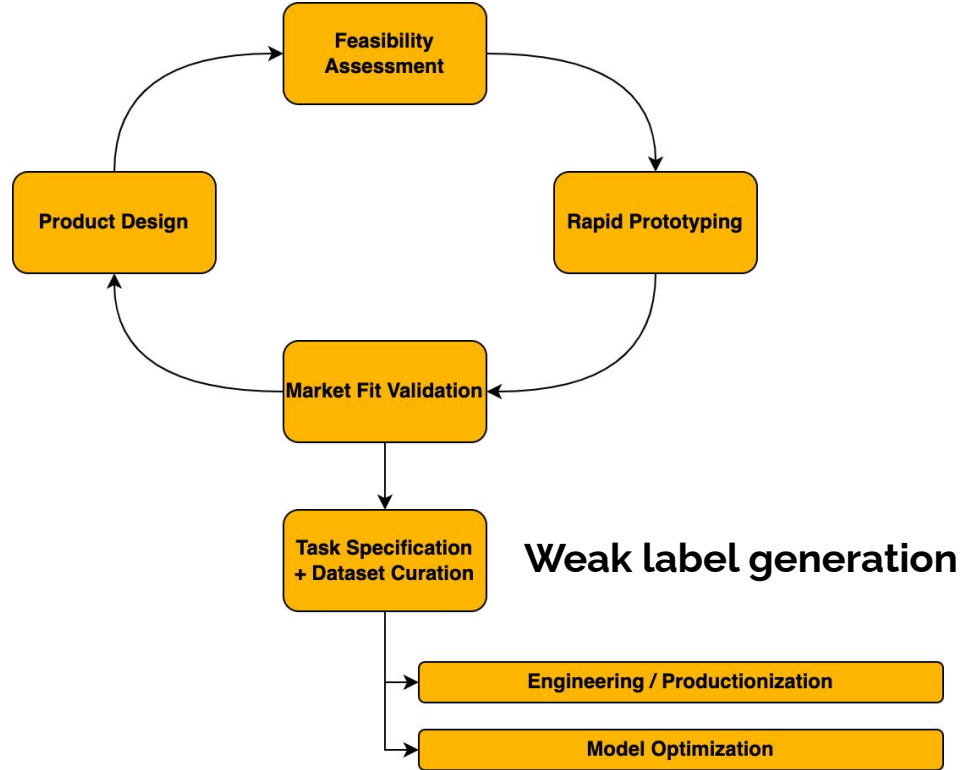


Figure 1: Models' classification accuracy (%) under different annotation and training costs (US Dollars).

"Which Model Shall I Choose? Cost/Quality Trade-offs for Text Classification Tasks" (Zong et al., 2023)





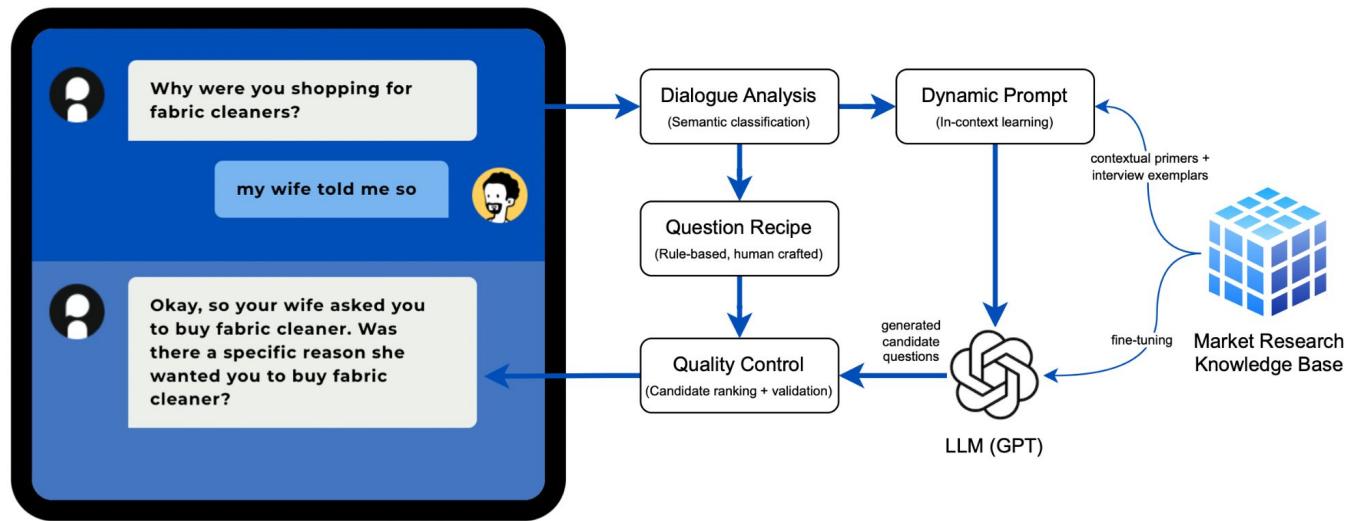


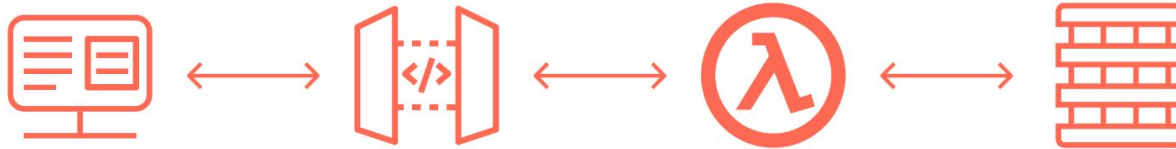
Figure 1: Processing flow of our SmartProbe, shown with an example input (prime question and response) and output (probing question). The Dialogue Analysis module first extracts structured textual data from previous conversations between an agent and the user. Our Dynamic Prompt module and LLM module, incorporating with Market Research Knowledge Base, then generate probing questions. For certain cases when the generated questions are not appropriate, we use the template questions from the Question Recipe module. Finally, the generated question will be checked by the Quality Control module to ensure it is of high quality and meets ethical standards.

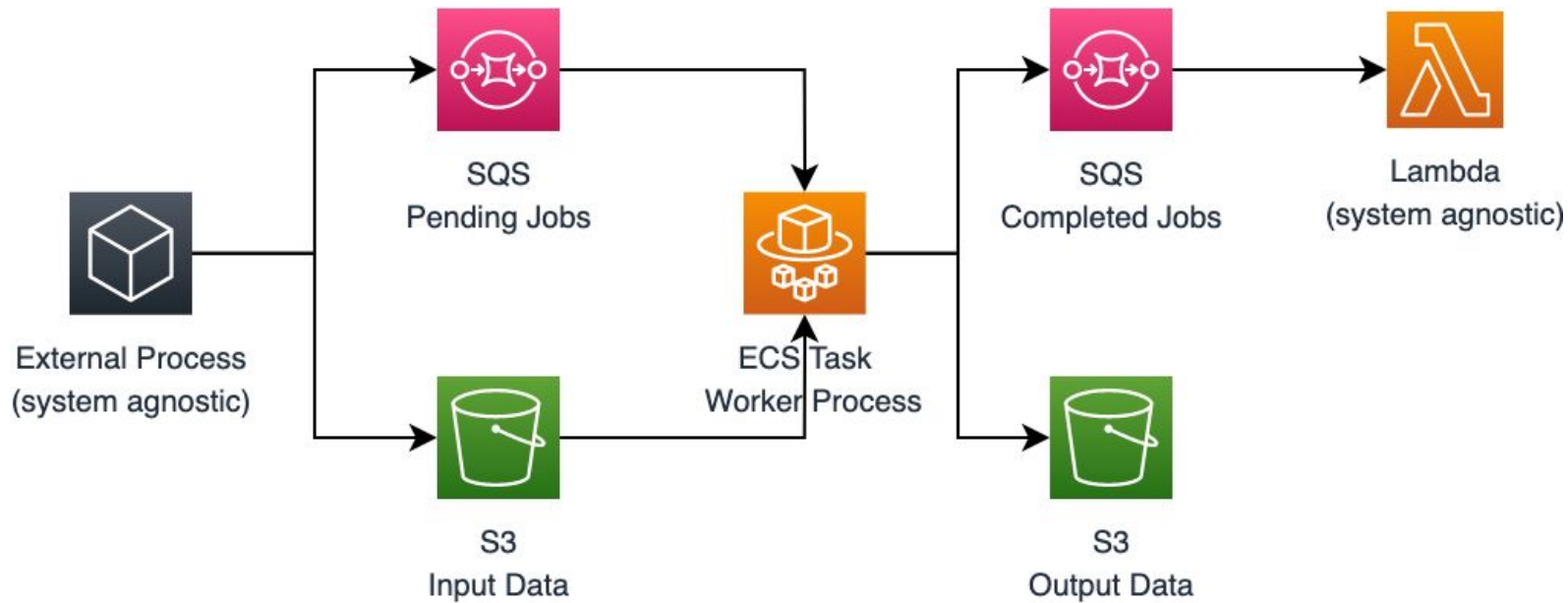
HTTP request

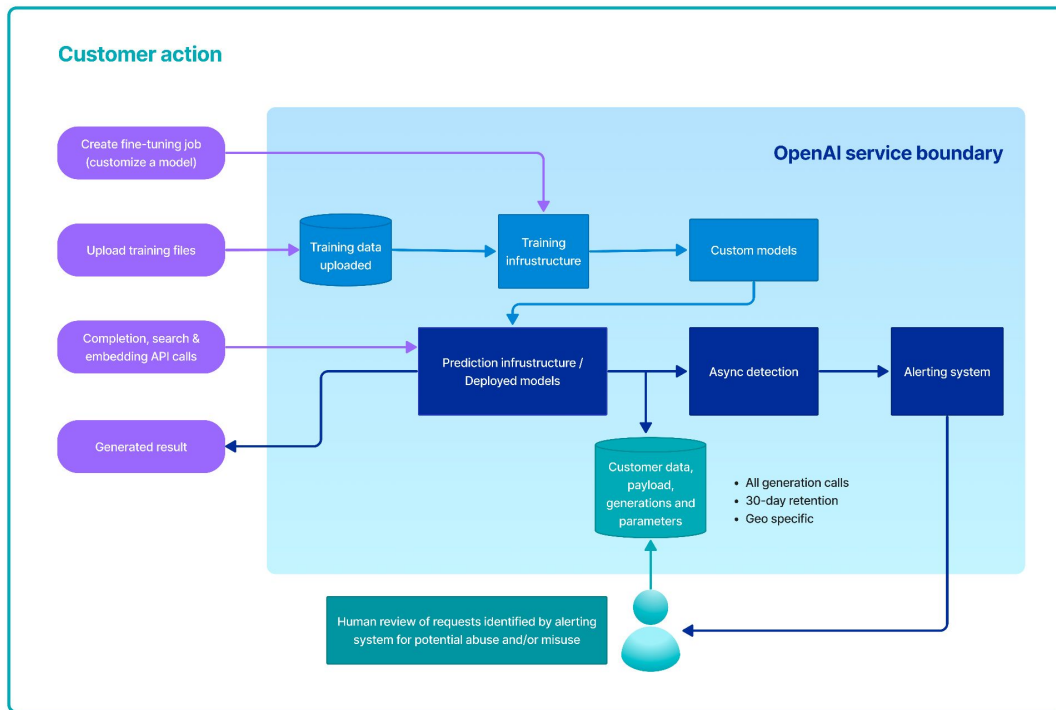
API Gateway

AWS Lambda

DynamoDB







Resources

- Infrastructure Credits
- Tax Credits
- Subsidized R&D
- Marketing Grants
- IP Support

Resources (in Canada)

- Infrastructure Credits = **Join an accelerator**
- Tax Credits = **SR&ED**
- Subsidized R&D = **Mitacs, IRAP, Innovation Clusters**
- Marketing Grants = **CDAP**
- IP Support = **IPON**

IP Strategy

- Applying LLMs to improve an existing capability may not be patentable...
...but solving an old problem in a brand new way might be!
- Human-in-the-loop interfaces which leverage LLMs to power solutions could warrant a design patent

Questions / Discussion?

josh@nexxt.in

<https://www.linkedin.com/in/josh-seltzer/>