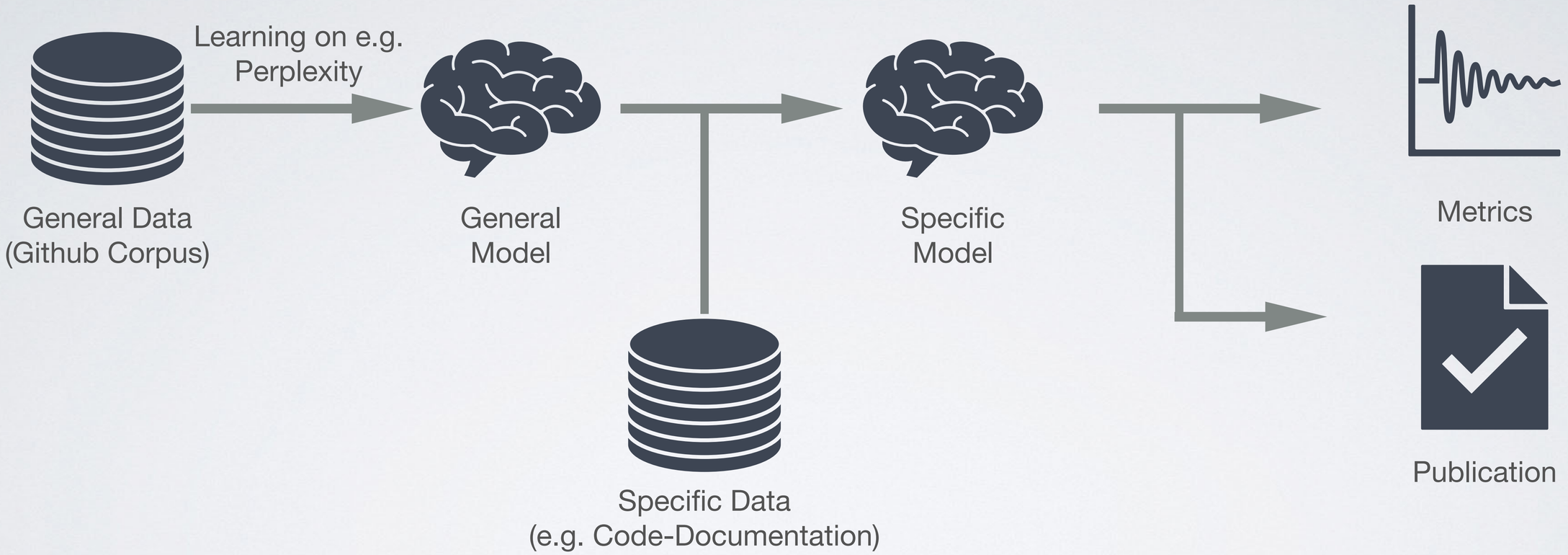
CISELab TU Delft

BLEU It All Away!

Leonhard Applis, TU Delft

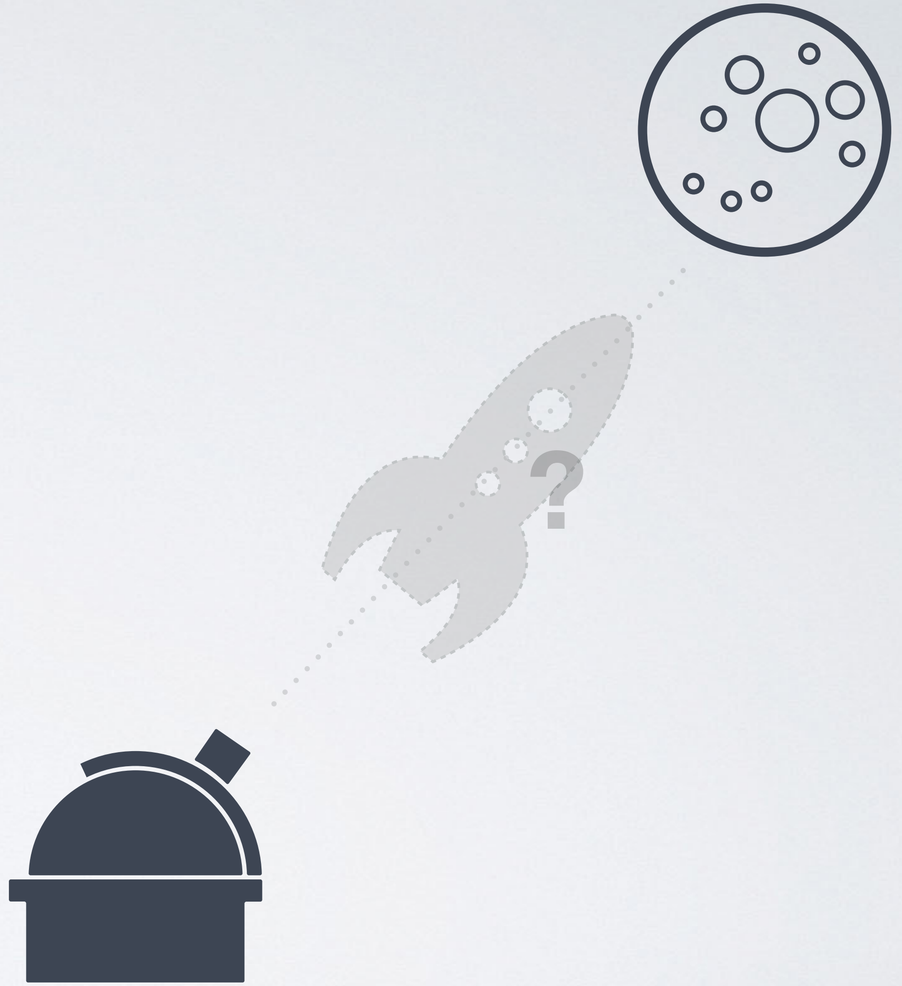
State of the Union

Current Pipeline



Current Approaches Are

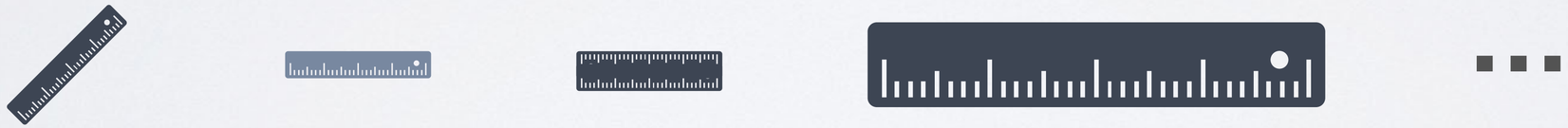
- Metric-Focussed
- Not used by developers
  (Method Name Prediction, anyone?)
- Tailor-made when actually used
  (Test-Generation @ Facebook)
- Made for academic publications



Current Metrics

- Broad Applicable Metric = Wide Adoption (F1 Score, BLEU)

- Thousands of strange Variations (CodeBLEU, BLEURT, MRR)



- **Acceptance of Metrics is seen as a proxy for acceptance of models!**

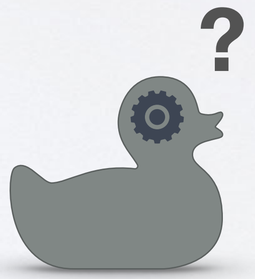
BLEU - Definition

$$BLEU = BP \times exp\left(\sum_{n=1}^{N} w_n \, log \, p_n\right)$$

- Brevity Penalty
- Supports Multiple References
- Supports Multiple N-Gram N's
- 2 "Studies" on Human Acceptance (For Translation)
- Handful Reports on Human Interaction

BLEU In Particular

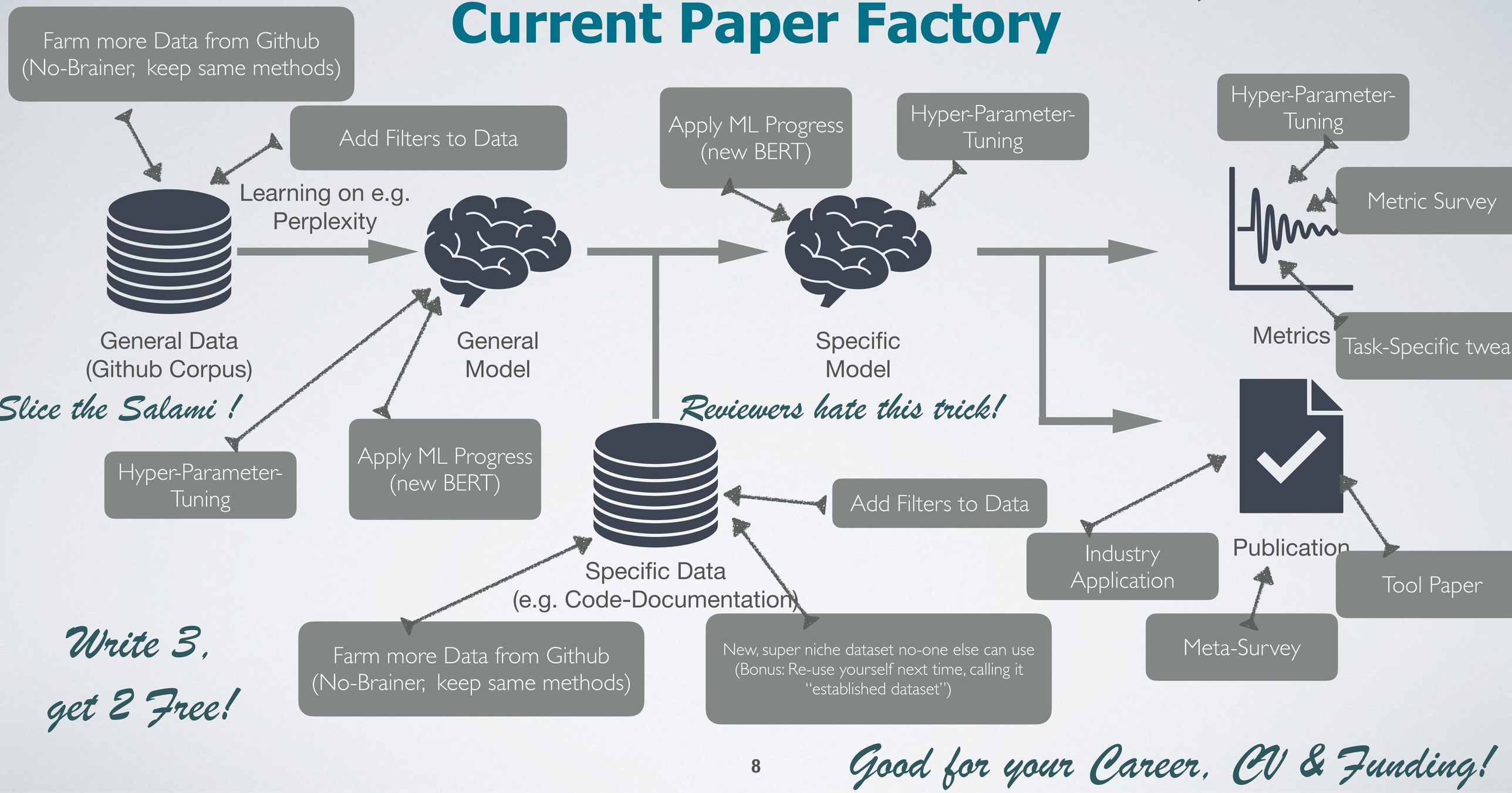
- [Ab]Used outside tested Domain
  (Generation Tasks, etc.)
- Noise-Baseline by Stop-Words
- SOTA Models
  worse than random texts [1]



[1] https://conf.researchr.org/details/icse-2022/icse-2022-posters/14/CrystalBLEU-Precisely-and-Efficiently-Measuring-the-Similarity-of-Code

Bring your Friends!

Current Paper Factory



Write 3,

get 2 Free!

Good for your Career, CV & Funding!

Science:
Incremental Vs. Malpractice

Reasonably sized contributions

Analysis of content, and of research itself

Useful for Domain, or at least for other researchers

Salami Slicing

Self-Citation

Unreproducible Experiments

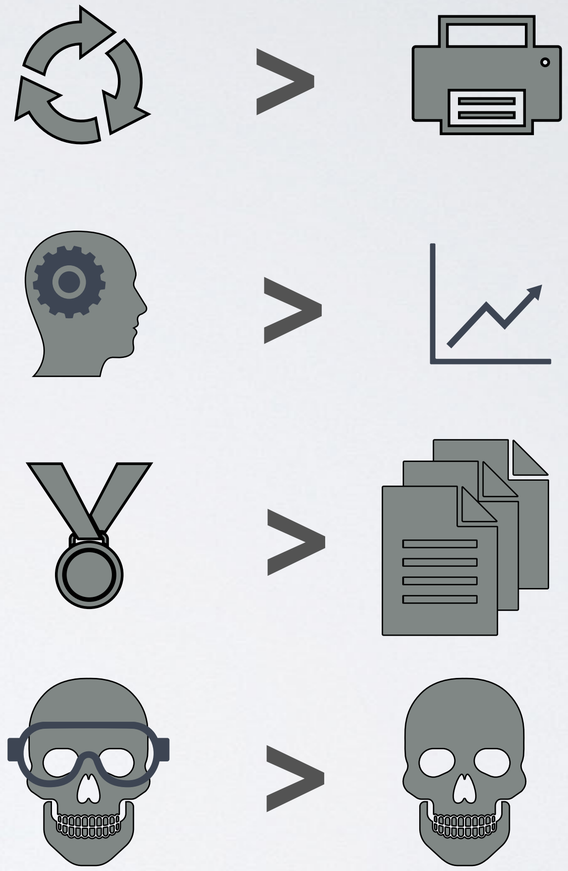
Intentionally not understandable

Intentionally not Exhaustive

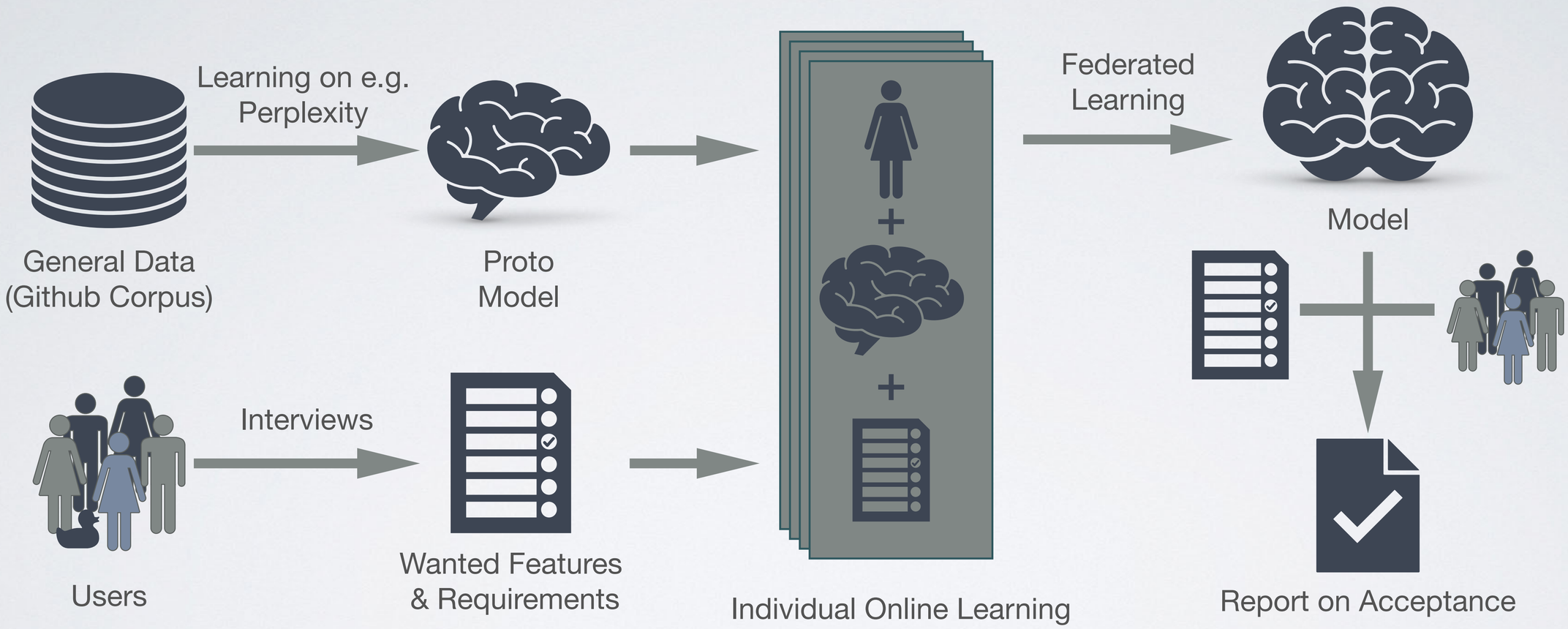
A Change of Hearts

What We Need

- Re-Usable (not only reproducible) Tools
- Real world Feedback (=Humans)
- High Quality, not High Quantity
  (For Data, Models, Papers, Studies)
- Tests



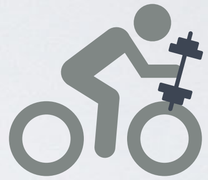
Proposed Pipeline



What We Have

- **Requirements Engineering:**
  Find out what really matters & prioritise

- **Reinforcement Learning:**
  Teach a model “on the fly”



- **Federated Learning:**
  Join efforts from all participants back into on model

- **Re-Usable & Safe Tools:**
  Self-Host with Containers, Contribute only what you want
  Provide life-examples on a webserver

Thank You!