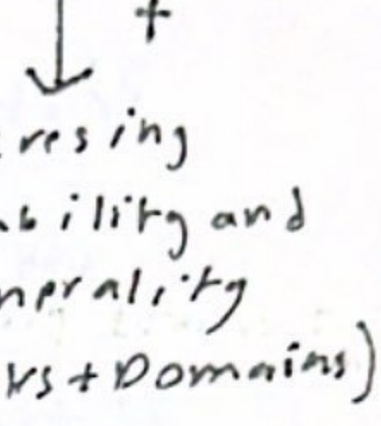
AGI / ASI

- ANI (narrow AI): Good at one thing i.e (captioning img, playing chess), what we have now.

- AGI (Artificial general intelligence): can understand learn and do most cognitive tasks that Humans can, across domains, with strong transfer + Autonomy.

- ASI (Artificial super intelligence): Far beyond human performance across virtually all domains like STEM.



- How to get AGI
  - Scale + Data + Compute
  - Better architectures
  - Reasoning
  - Agents
  - COT
  - RAG
  - Multimodal
  - Distillation + reflection

- Why AGI is hard
  - Alignment and Safety
  - Deceiving
  - Feedback / planning

- ASI = Beyond the best human experts at Everything but Get AGI First
- AGI and ASI can do thing independently with Little/ to no human Feedback

NOTE: AGI ≠ consciousness → not req

Bigger ≠ AGI → Scaling helps but reasoning memorize matters too

Boxing ≠ Enough → isolation helps but is not enough

Alignment by making it nice ≠ work! too probabilistic
cannot control output just train that

Safety: RLHF, REACT