

On the Measure of Intelligence

-Francois Chollet

Stages of Paper:



Context and History about the
measure of intelligence



New Perspective by Francois
Chollet



Benchmark Proposal: ARC

Defining Intelligence:



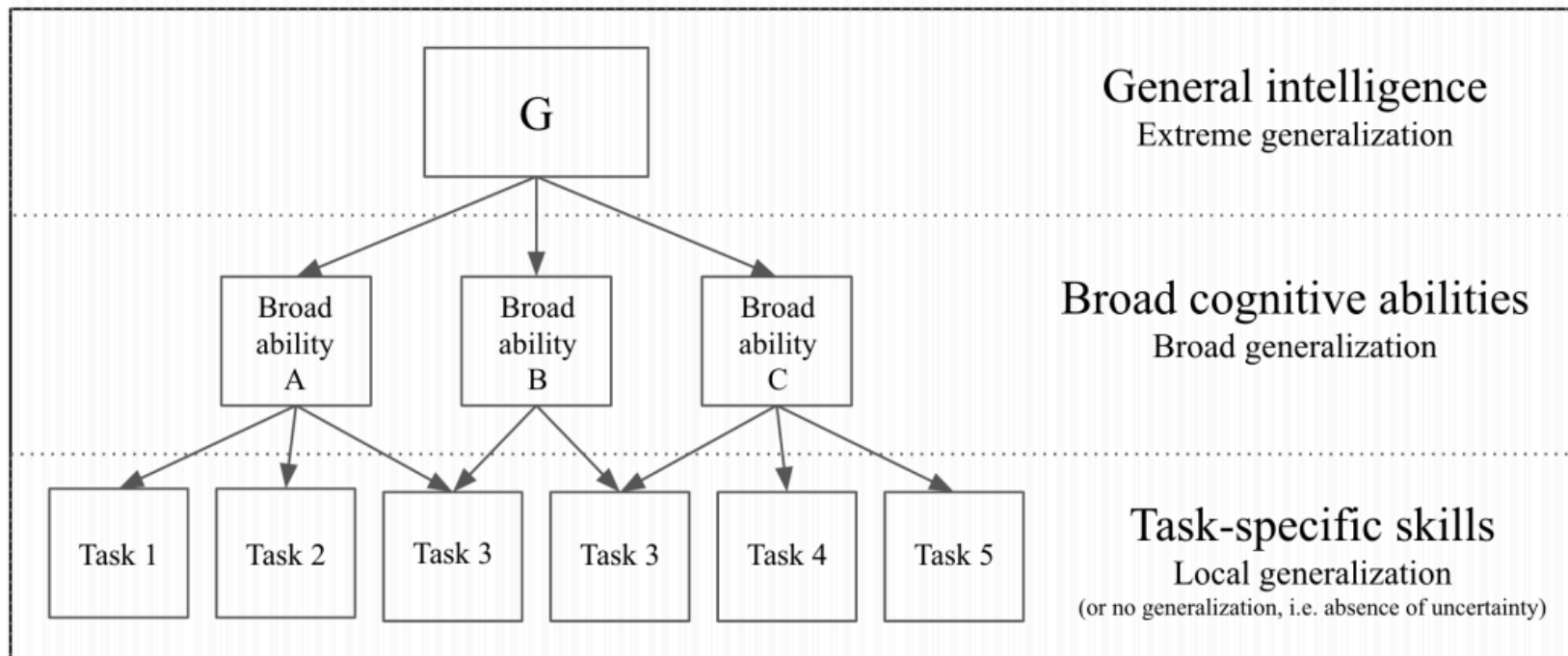
Collection of task specific skills



General learning ability

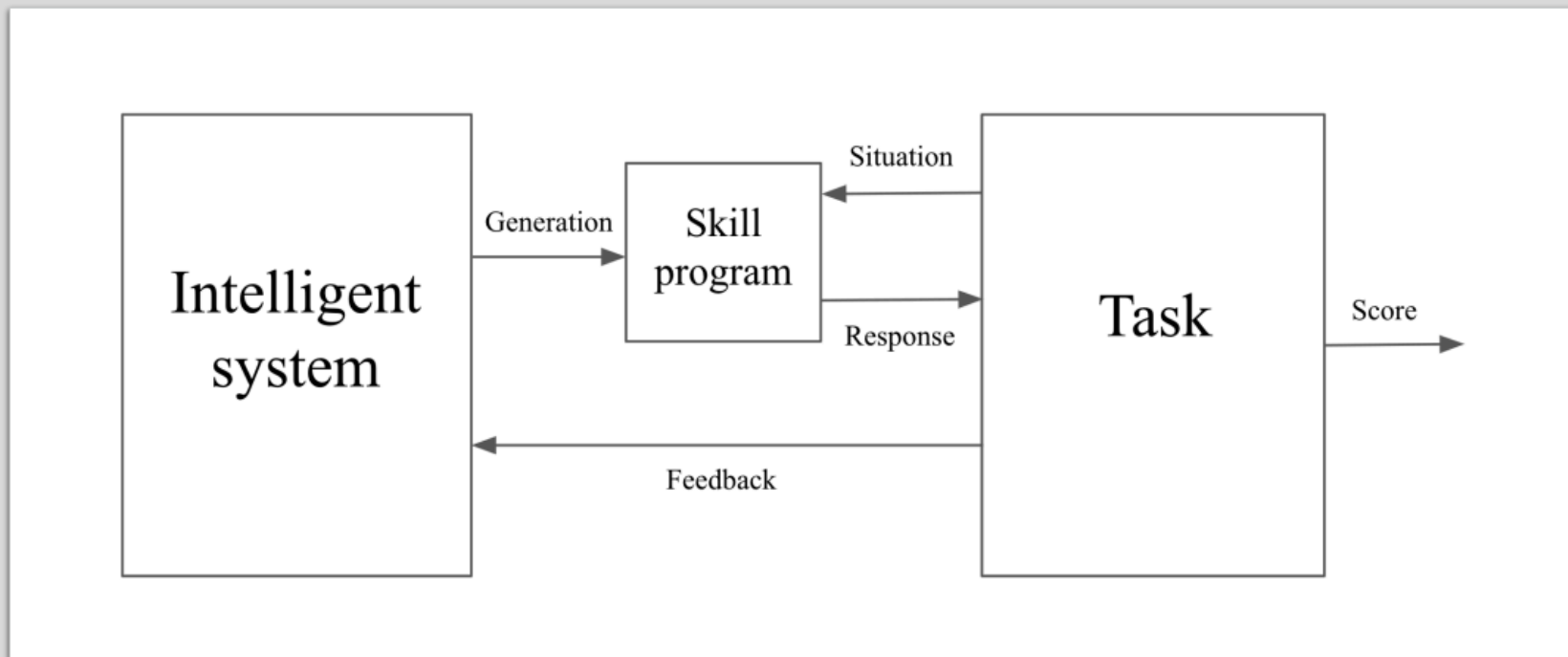
AI Evaluation

- Skill Based Evaluation (Single Task)
 - Alex Net – Image Net
 - Alpha Go
 - GPT-3 – Text Generation
- Generalization(Multiple Tasks)
 - Agent 57
 - Types Of Generalization:
 - Absence – Programs Like Searching and Sorting
 - Local – When we input the Dense Data set the agent gets an idea
 - Broad – Agent has never seen this data nor trained and has to perform on it
 - Extreme – Good at almost every task



Position Of The Problem:

An intelligent system generates a skill program to interact with a task



New Perspective by Francois Chollet

- Chollet Definition Of Intelligence:

The Intelligence of a system is a measure of its skill acquisition efficiency over a scope of tasks, with respect to priors, experience and generalization difficulty.

- Priors and Experience

“If there exists a task with HIGH GENERALIZATION DIFFICULTY and an AGENT can solve it with LESS PRIORS or with LESS EXPERIENCE, then it is INTELLIGENT”

EXPECTATIONS [SKILL GENERALIZATION/PRIORS+EXPERIENCE]



Generalization, Priors, Efficiency:

$$GD_{T,C}^{\theta} = \frac{H(Sol_T^{\theta} | TrainSol_{T,C}^{opt})}{H(Sol_T^{\theta})}$$

$$P_{IS,T}^{\theta} = \frac{H(Sol_T^{\theta}) - H(Sol_T^{\theta} | IS_{t=0})}{H(Sol_T^{\theta})}$$

$$E_{IS,T,t}^{\theta} = H(Sol_T^{\theta} | IS_t) - H(Sol_T^{\theta} | IS_t, data_t)$$

$$GD_{IS,T,C}^{\theta} = \frac{H(Sol_T^{\theta} | TrainSol_{T,C}^{opt}, IS_{t=0})}{H(Sol_T^{\theta})}$$

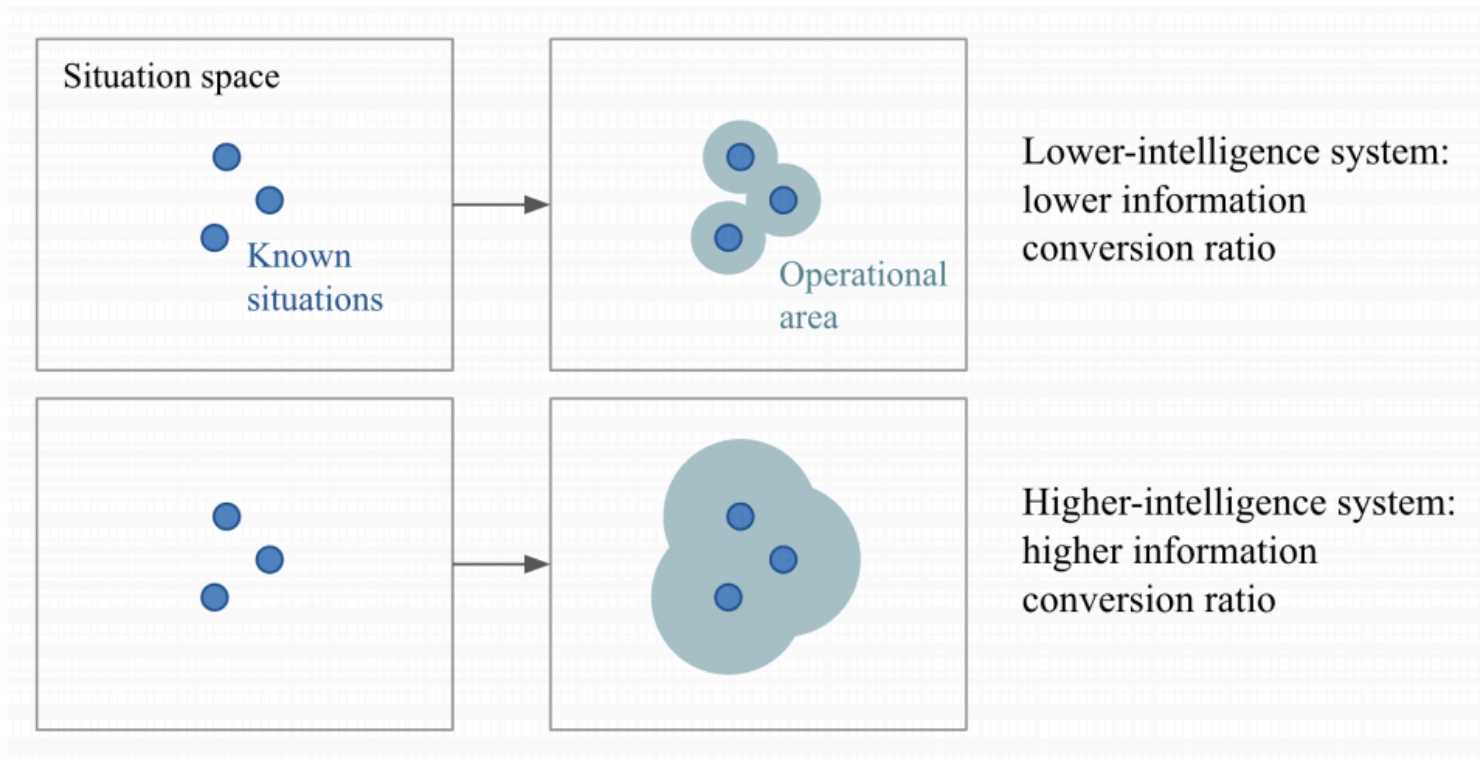
$$E_{IS,T,C}^{\theta} = \frac{1}{H(Sol_T^{\theta})} \sum_t E_{IS,T,t}^{\theta}$$

Intelligence of system IS over $scope$ (sufficient case):

$$I_{IS,scope}^{\theta_T} = Avg_{T \in scope} \left[\omega_T \cdot \theta_T \sum_{C \in Cur_T^{\theta_T}} \left[P_C \cdot \frac{GD_{IS,T,C}^{\theta_T}}{P_{IS,T}^{\theta_T} + E_{IS,T,C}^{\theta_T}} \right] \right]$$

Intelligence of system IS over $scope$ (optimal case):

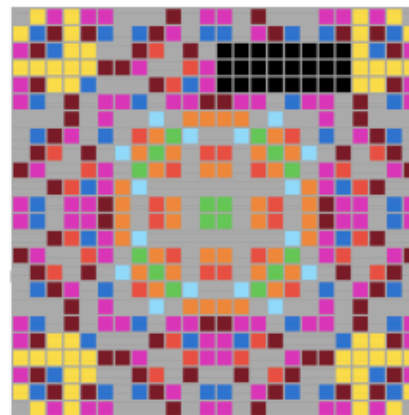
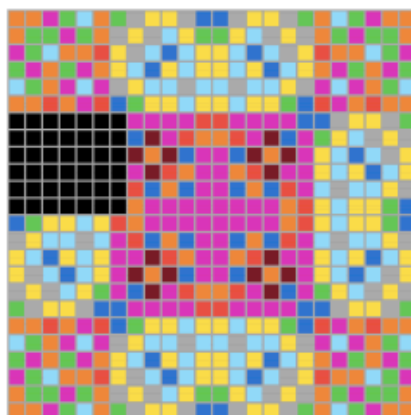
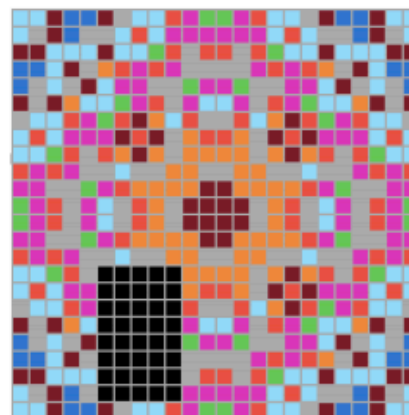
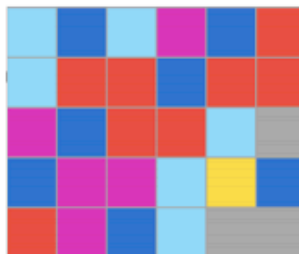
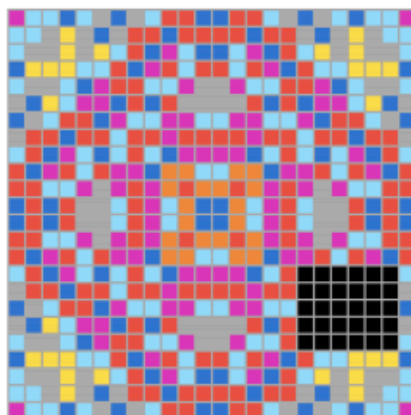
$$I_{IS,scope}^{opt} = Avg_{T \in scope} \left[\omega_{T,\Theta} \cdot \Theta \sum_{C \in Cur_T^{opt}} \left[P_C \cdot \frac{GD_{IS,T,C}^{\Theta}}{P_{IS,T}^{\Theta} + E_{IS,T,C}^{\Theta}} \right] \right]$$



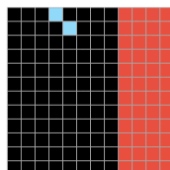
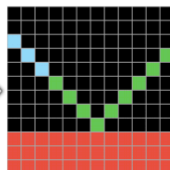
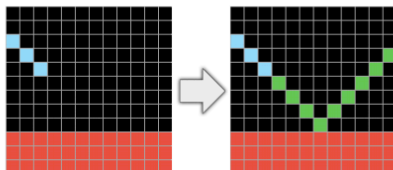
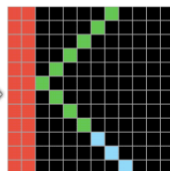
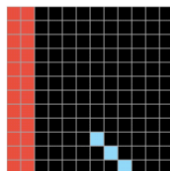
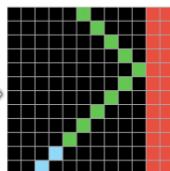
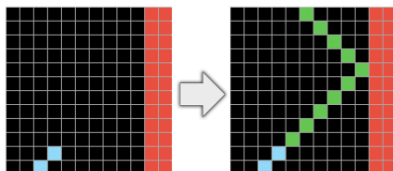
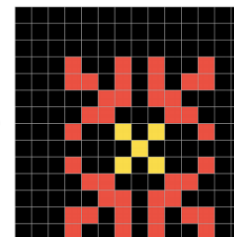
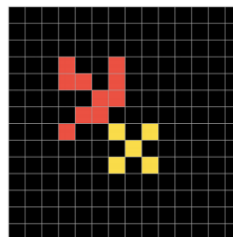
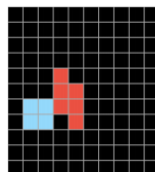
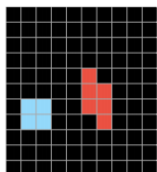
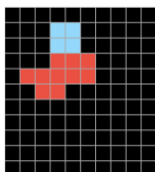
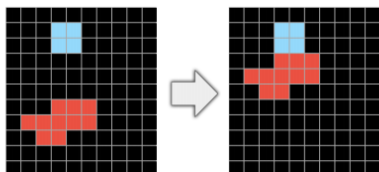
Benchmark Proposal: ARC dataset

- ARC (abstraction and reasoning corpus)
- *ARC can be seen as a general artificial intelligence benchmark, as a program synthesis benchmark, or as a psychometric intelligence test. It is targeted at both humans and artificially intelligent systems that aim at emulating a human-like form of general fluid intelligence."*

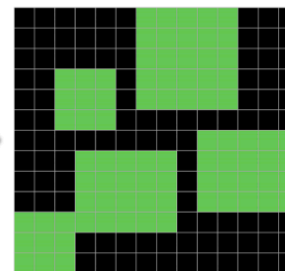
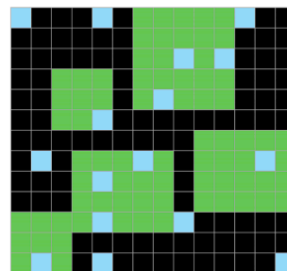
(<https://github.com/fchollet/ARC>)



?



?



Weakness

- Generalization is not quantified
- Test Validity is not established
- Data size and diversity might be limited
- Knowledge on priors may not of be well crafted