A Mathematical Model of AutoResearch

Inspired by `karpathy/autoresearch` — March 2026

Definition. *An* ***AutoResearch process*** *is a discrete-time stochastic optimization equivalent to a* ***(1+1) Evolution Strategy*** *with elitist selection: a single solution is maintained, a single variant is proposed at each step, and the better of the two survives. The process is fully determined by* ***3 parameters*** $(\sigma_0,\ \gamma,\ \mu)$ *and operates over a fixed budget of* N *experiments, each lasting* τ *seconds.*

1. Update rule. Let Θ be the space of code configurations and $f : \Theta \to \mathbb{R}^+$ the objective (`val_bpb`, lower is better). At each step t the agent proposes a variant θ'_t:

$$\theta^*_{t+1} = \begin{cases} \theta'_t & \text{if } f(\theta'_t) < f(\theta^*_t) \quad (\text{KEEP}) \\ \theta^*_t & \text{otherwise} \quad (\text{DISCARD}) \end{cases}$$

2. Perturbation model. The agent's modification is modeled as a stochastic perturbation in metric space:

$$f(\theta'_t) = f(\theta^*_t) + \delta_t\,, \qquad \delta_t \sim \mathcal{N}(\mu,\ \sigma_t^2)\,, \qquad \sigma_t = \sigma_0 \cdot \gamma^{\,t}$$

3. The three parameters.

Symbol	Name	Range	Meaning
σ_0	Initial step size	$\sigma_0 > 0$	How bold the agent's changes are at the start
γ	Decay rate	$\gamma \in (0,\ 1)$	How fast exploration narrows (like annealing temperature)
μ	Bias	$\mu \leq 0$	Agent skill: $\mu = 0$ is random, $\mu < 0$ biases toward improvement

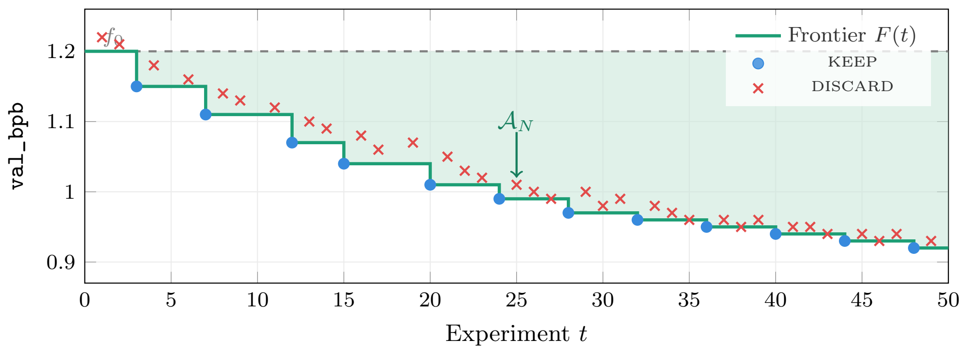
The keep probability follows from the perturbation distribution: $\Pr[\text{KEEP}] = \Phi\big(-\mu/\sigma_t\big)$, which equals $\frac{1}{2}$ when $\mu = 0$ (random agent) and grows as $|\mu|$ increases (skilled agent).

4. Frontier and cumulative gain area. The frontier $F(t) = \min_{s \leq t} f(\theta^*_s)$ is non-increasing by construction. The **cumulative gain area** measures total improvement over the baseline f_0:

$$\boxed{\mathcal{A}_N = \sum_{t=1}^{N} \big[f_0 - F(t)\big]}$$

Maximizing $\mathcal{A}_N$ rewards both *fast convergence* (the frontier drops early) and *final depth* (the frontier reaches a low value). In the symmetric case ($\mu = 0$):

$$\mathbb{E}[\mathcal{A}_N] \approx \frac{\sigma_0}{\sqrt{2\pi}\,(1-\gamma)} \left[N - \gamma\,\frac{1-\gamma^N}{1-\gamma} \right]$$



5. Trade-offs. High σ_0: large jumps, many discards. Low σ_0: small gains, many keeps. $\gamma \to 1$: explores longer. Low γ: converges early, risks local optima.

6. Extensions. **Multi-agent** (k parallel agents): $\Pr[\text{at least 1 keep}] = 1 - (1-p)^k$. **Meta-optimization**: $\texttt{program}^* = \arg\max_{\texttt{prog}} \mathbb{E}[\mathcal{A}_N \mid \texttt{prog}]$.