Fil ve Japon Balığı modeli: kurulum ve kullanım

Yapay zeka, dışarıdan geri bildirim almadan kendi hatasını bulmakta zorlanıyor. Fil ve Japon Balığı (Elephant Goldfish Model, kısaca EGM) bunu iki rolle çözüyor: her şeyi bilen bir "fil" planı yazıyor, sohbeti hiç görmemiş "japon balıkları" o planı inceliyor. Kod, sen planı onaylamadan yazılmıyor. Ücretsiz ve açık kaynak; Google'ın GitHub hesabında yayınlandı. Bilgiler 7 Kasım 2026'da deponun kendi belgelerinden kontrol edildi.

0 Kim yaptı, ne durumda?

APACHE-2.0 LİSANSI

`github.com/` **`google/elephant-goldfish-model`** ★ 13.527 yıldız

Yöntem Google'da Distinguished Engineer olan Dave Rensin'e ait; ekibi içeride kullandıkları eklentiyi 21 Ekim 2026'da açık kaynak yaptı. Depo Google'ın resmi GitHub hesabında ama README açıkça şunu söylüyor: **resmi destekli bir Google ürünü değil**. Projenin sitesi (egm.dev) "binlerce Google mühendisinin kullandığını" yazıyor; bu sitenin kendi iddiası, sayı ya da yöntem verilmiyor. Son sürüm v0.1.1 (6 Kasım 2026).

1 Nasıl çalışır: üç rol

FİL
Senin ana oturumun
Konuşmanın tamamını ve verdiğin belgeleri bilir. Kod yazmadan önce tasarım taslağını (plan) yazar.

JAPON BALIKLARI
Temiz oturumlu ajanlar
Sohbeti görmezler; yalnız taslağı ve ilgili kodu okurlar. Fikir aşamasında 7 farklı mercekle, incelemede 5 aşamada toplam 10 kez bakarlar.

SEN
Son karar
Bulguları seninle birlikte çözerler, planı sen onaylarsın. EGM kod yazmaz: onaydan sonra ajanına "bu tasarıma göre uygula" dersin.

2 Kurulum

CLAUDE CODE 2.1.271 VE ÜSTÜ

TERMİNALDE

```
claude plugin marketplace add google/elephant-goldfish-model
claude plugin install egm@egm
```

EKLENTİ LİSTEDE GÖRÜNMEZSE

```
claude plugin enable egm@egm
```

Kurduktan sonra **yeni bir oturum aç**. Ajanın alt ajan (subagent) açabilmesi gerekiyor; açamıyorsa inceleme durur. Claude Code dışında Codex, ChatGPT masaüstü ve Antigravity'de de çalışıyor; kurulumlar deponun docs/harnesses.md dosyasında.

3 Komutlar

CLAUDE CODE'DA /egm: ÖNEKİYLE

`/egm:egm-brainstorm`	Önce netleştirici sorular sorar, sonra balıklar fikri sınar. Çıktı: concepts_brief.md
`/egm:egm-draft`	Tasarım taslağını yazar (design_draft.md). Onayladığın tasarım docs/designs/ altına kaydedilir.
`/egm:egm-review`	5 aşamalı inceleme: toplam 10 inceleme, sonunda bir özet.
`/egm:egm-resolve`	İnceleme bulgularını seninle tek tek çözer.
`/egm:egm-drift-report`	Depodaki belgeleri güncel kodla karşılaştırır. Kod yazıldıktan sonra planla kod arasındaki farkı görürsün.
`/egm:egm-bootstrap`	Var olan bir kod tabanı için belgeleri oluşturur ya da günceller.

4 İlk deneme: adım adım

1. **Büyük bir iş seç.** Ödeme akışı, bildirim sistemi, yeni bir modül gibi. Tek satırlık düzeltme için kurma, değmez.
2. **Fikri sınat.** Aşağıdaki gibi başla; balıkların sorularını cevapla.

```
/egm:egm-brainstorm Siparişte ödeme alınca müşteriye fatura e-postası gönderen bir akış ekle.
Mevcut ödeme kodu src/odeme klasöründe. Belgeleri Türkçe yaz.
```

3. **Taslağı yazdır:** /egm:egm-draft. Her bölümü oku, gerekiyorsa düzelt, sonra onay ver.
4. **İncelet ve çöz:** /egm:egm-review, ardından /egm:egm-resolve. "Ya işlem yarıda kesilirse?" gibi soruların hepsine bir cevap bul.
5. **Onayla ve uygulat:** "docs/designs/ altındaki tasarıma göre uygula" de. Kod bitince /egm:egm-drift-report ile kodun plana uyup uymadığına bak.

Neden işe yarıyor?

ARAŞTIRMA

- **Huang ve ark., ICLR 2024 (Google DeepMind):** büyük dil modelleri dışarıdan geri bildirim olmadan kendi akıl yürütme hatalarını düzeltmekte zorlanıyor; bazen sonuç daha da kötüleşiyor.
- **Tyen ve ark., ACL 2024 Findings:** asıl zorluk hatayı **bulmak**. Hatanın yeri söylenince modeller onu iyi düzeltebiliyor.
- **Google'ın tasarım belgesi kültürü:** Google'da mühendisler kod yazmadan önce bir tasarım belgesi yazıp incelemeye açıyor (Malte Ubl, "Design Docs at Google"). EGM bu alışkanlığı ajanlara taşıyor.

Dikkat edilecekler

- **Token tasarrufu değil, ek harcama:** bir plan için 7 ila 10'dan fazla ajan çalışıyor; Claude planının kullanım limiti daha hızlı dolar. Anthropic'in kendi ölçümünde çok ajanlı bir sistem, normal sohbetin yaklaşık 15 katı token harcadı.
- **Kodu değil planı inceler:** README'ye göre bu inceleme testlerin ve insan incelemesinin yerini tutmaz. Testleri yine yaz, kodu yine oku.
- **Küçük işte değmeyebilir:** projenin SSS'si küçük değişikliklerde bu akışın karşılığını vermeyebileceğini söylüyor.
- **Fil yalnız bildiğini bilir:** "her şeyi bilir" dediğimiz şey, oturumdaki konuşma ve ona verdiğin belgeler. Önemli bir kısıt varsa açıkça yaz.

Türkiye için notlar

- **Eklenti ücretsiz, Claude Code ücretli:** Claude Code ücretsiz planda yok; en uygunu Pro, aylık 20 dolar (yıllık ödemede aylık 17 dolar). Claude, Anthropic'in desteklenen ülkeler listesinde Türkiye'yi kapsıyor.
- **Plan Türkçe olabilir:** komutlar İngilizce ama isteğini Türkçe yazabilir, belgelerin Türkçe yazılmasını isteyebilirsin. Ekip arkadaşların okuyacaksa bu işini kolaylaştırır.

Ücretsiz kılavuz · futurewithserdar.com · @future_with_serdar · **FİL** yazanlara özel hazırlandı.
Videodaki fatura örneği ve ekranlar temsili. Yıldız sayısı GitHub, 7 Kasım 2026.
Kaynaklar: github.com/google/elephant-goldfish-model · egm.dev · arxiv.org/abs/2310.01798 (Huang ve ark.) · arxiv.org/abs/2311.08516 (Tyen ve ark.) · anthropic.com (çok ajanlı sistem) · code.claude.com (alt ajanlar) · industrialempathy.com (Design Docs at Google) · claude.com/pricing