

## Architecture Vision: From Idea to High-Load System

Author: Victor Pisarevsky, CTO DIVO

Topic: How we design Enterprise+AI solutions under uncertainty.

### 1. Introduction: Why GPT Wrappers Die in Production

Most AI projects start with excitement: "Look, ChatGPT wrote an email!". But when such a prototype faces reality, it breaks.

#### Typical Problems:

- **Ambiguity:** Lack of clear and regulated processes that are "simple" enough to transfer into an IT system.
- **Latency:** The user will not wait 15 seconds for an answer.
- **Data Security:** There are categories of data that have no right to leave the company's information perimeter.
- **Legacy:** The solution must interact with a dozen systems, half of which lack a clear API and/or run on ancient protocols.
- **Hallucinations:** The operation of neural networks is non-deterministic; identical inputs can generate different outputs.

We at DIVO do not build "chatbots" or "magic assistants". We build **Intelligent Systems** – engineering solutions where the neural network is just one of the modules, wrapped in rigid logic and secure infrastructure.

---

### 2. Our Approach: The "DIVO Framework"

We don't start by writing code. We start with architecture.

Our process (Discovery Phase) allows turning an abstract idea into a clear action plan in 2-3 weeks.

#### Step 1. Landscape Audit and Process Synthesis

- **Essence:** Turning business "pain" into a technical strategy.  
We do not require ready-made technical specifications (ToR), as this is impossible in AI projects. Instead, we study the current systems (Legacy) and business goals.
- **Action:** We collaboratively design the target process ("To-Be"), answering two questions:
  1. Exactly what routine are we offloading to AI?
  2. How do we fit this into the current IT landscape without disrupting established business processes?
- **Result:** A roadmap that takes real system constraints into account.

#### Step 2. Feasibility Check (PoC)

- **Essence:** Cheap hypothesis testing before expensive development begins.  
Before writing complex code and touching databases, we verify if AI is capable of solving

- the task in principle.
- **Action:** We take data samples (or generate similar ones) and run them through models in an isolated environment.
- **Result:** Confirmation that the technology works, or an expectation adjustment (pivot) without wasting the budget on full-scale development.

### Step 3. Data Security Engineering

- Essence: Designing an architecture where data leakage is impossible. We create a system that accounts for the new challenges created by neural network-based modules in the context of cybersecurity.
- **Action:** All data the neural networks will work with is labeled by criticality. Depending on the result, we select the appropriate method of interaction (public services/private cloud/isolated module). Input and output data from such modules must comply with pre-approved JSON/XML schemas. AI agents are granted strictly those rights and access levels necessary for their work.
- **Result:** The AI module performs the assigned task and is incapable of any other actions, preventing its use by malicious actors.

### Step 4. High-Load Handling

- Essence: Preparing the solution for scaling under real load. AI may work slowly, and Legacy systems may have performance limitations, but the system must be able to handle the entire user flow.
- **Action:** System component interaction is built asynchronously; calls to Legacy components occur strictly through buffer adapters with regulated throughput.
- **Result:** We protect old systems from overloads caused by AI modules, ensuring the stability of the entire perimeter even at peak loads.

### Step 5. Evolutionary Launch and Fine-Tuning

- Essence: Smooth entry into operation. We do not switch on automation with a "big bang" for everyone at once.
- **Action:** The launch proceeds through stages:
  - **Shadow Mode:** AI works "idly"; we compare its decisions with human decisions.
  - **Pilot:** AI helps a limited group of employees.
  - **Prod:** The system rolls out to everyone, continuously learning from user feedback.
- **Result:** Predictable quality and no risk of business stoppage.

## 3. How We Work: From Audit to Launch

We offer a transparent process where you pay for the stage result, not for "development hours".

1. **Technical Audit (2 weeks):** Steps 1 and 2. **Result:** Architecture Vision Document (AVD)

- and budget estimation.
2. **PoC (Proof of Concept) (4 weeks):** Step 3. We assemble a prototype on real data to check answer quality. If the result is satisfactory, we move to production.
  3. **Production (3-6 months):** Steps 4 and 5. Microservice development, CI/CD setup, load testing, monitoring implementation.
- 

## Conclusion

AI is not magic. It is a technology requiring engineering discipline.

If you are ready to build a system that will work for years and scale, we are ready to become your technology partner.

**DIVO Engineering Team**