

# NabavkiData

## EIC Accelerator Step 1 - Form Answers

Complete Short Proposal Application

COMPANY	COUNTRY	TRL	FUNDING REQUEST
<b>NabavkiData d.o.o.e.l.</b>	<b>North Macedonia</b>	<b>7</b>	<b>€2,500,000</b>

---

## SECTION 1: EXCELLENCE

### 1.1 Innovation - What is the innovation?

NabavkiData is an AI-powered public procurement intelligence platform that transforms unstructured government data into actionable insights for businesses, anti-corruption organizations, and citizens across the Western Balkans.

#### Core Innovation - Three Interconnected Breakthroughs:

##### 1. Cyrillic Deep OCR Pipeline (Our Primary Technical Moat)

Off-the-shelf OCR solutions achieve only 40-60% accuracy on degraded Cyrillic government scans. We developed a custom-trained pipeline achieving **94% accuracy**:

- Pre-processing optimized for low-quality government scans
- Custom models trained on 2+ years of real procurement documents
- Post-processing with domain-specific entity recognition
- Automatic extraction of structured fields from unstructured text

##### 2. RAG-Powered Semantic Search

- 450,000+ document embeddings for semantic similarity
- Bilingual query processing (automatic Latin ↔ Cyrillic transliteration)
- Retrieval-Augmented Generation for natural language Q&A
- Context-aware responses with source citations

##### 3. ML-Based Corruption Detection (Novel Application)

- **Anomaly Detection:** Statistical flagging of unusual pricing, timing patterns

- **Graph Neural Networks:** Network analysis to identify collusion rings
- **Temporal Analysis:** Pattern recognition for suspicious seasonal publication
- **Red Flag Scoring:** Automated risk assessment per tender and entity

## 1.2 Beyond State of the Art

Dimension	Current SOTA	NabavkiData Advancement
OCR Accuracy	40-60% on Cyrillic	94% accuracy (2x improvement)
Search	Keyword matching	Semantic understanding
Language	Single language	Bilingual (Latin/Cyrillic)
Coverage	Single country or EU only	Multi-country Western Balkans
History	Current year	16 years (2008-2024)
Freshness	Weekly/monthly	3-hour sync cycle
Corruption Detection	Manual investigation	Automated ML flagging

## 1.3 Technology Readiness Level

**Current TRL: 7** (System prototype demonstrated in operational environment)



Component	TRL	Evidence
Web scraping infrastructure	8	Production, 3-hour sync cycle

Cyrillic OCR pipeline	7	94% accuracy, 32K+ documents processed
Embedding generation	7	450,000+ embeddings in production
RAG search system	7	Live with bilingual support
Corruption detection ML	5-6	Algorithms designed, initial validation

## 1.4 Intellectual Property Strategy

IP Asset	Status	Protection Strategy
Cyrillic OCR Models	Developed	Trade secret (architecture + training data)
Corruption Detection Algorithms	In development	Patent application planned Q2 2025
Bilingual NLP Pipeline	Developed	Trade secret
Training Data	2+ years accumulated	Trade secret (key competitive moat)
NabavkiData Brand	Registered	EU trademark application filed

**Freedom to Operate:** No blocking patents identified. Built on open-source foundations with no licensing constraints. Preliminary FTO review completed.

## SECTION 2: IMPACT

### 2.1 Market Opportunity - TAM/SAM/SOM

**€850M**

TAM (EU + Candidates)

**€120M**

SAM (Cyrillic Markets)

**€15M**

SOM (2028 Target)

#### Market Timing - Why Now:

Driver	Trend	Impact
EU Enlargement Push	Western Balkans priority post-Ukraine	Chapter 23 compliance mandatory
AI Breakthrough	LLMs, embeddings now viable	Semantic search finally possible
Digital Transformation	Post-COVID government digitization	More data available online
Anti-Corruption Funding	€2B+ EU investment in rule-of-law	Willing buyers for our tools

### 2.2 Scalability

Phase	Countries	Timeline	Users
Complete	North Macedonia	Done	5,000
Phase 2	Kosovo, Albania	2025	20,000
Phase 3	Serbia, Montenegro	2026	60,000

Phase 4	Bosnia, Croatia, Bulgaria	2027+	120,000
---------	---------------------------	-------	---------

## 2.3 EU Strategic Alignment

<p><b>EU Priorities</b></p> <ul style="list-style-type: none"> <li>• <b>EU Enlargement:</b> Chapter 23 compliance support</li> <li>• <b>European Green Deal:</b> Green procurement oversight</li> <li>• <b>Digital Decade 2030:</b> AI-powered transparency</li> <li>• <b>Anti-Corruption Strategy:</b> Technology infrastructure</li> </ul>	<p><b>UN SDG Alignment</b></p> <ul style="list-style-type: none"> <li>• <b>SDG 16:</b> Peace, Justice, Strong Institutions</li> <li>• <b>SDG 8:</b> Decent Work &amp; Economic Growth</li> <li>• <b>SDG 9:</b> Industry, Innovation, Infrastructure</li> <li>• <b>SDG 10:</b> Reduced Inequalities</li> </ul>
--	---

### Measurable Impact Targets (by 2028):

Metric	Target	Measurement
Corruption cases identified	50+	Referrals to authorities
SME bid success improvement	+25%	Before/after analysis
Procurement savings identified	€50M+	Price anomaly detection
Documents made searchable	2M+	Platform metrics
Countries covered	5+	Platform deployment

## 2.4 Competitive Advantage

Barrier	Our Position	Time to Replicate
Training Data	2+ years of Cyrillic documents	2-3 years minimum
OCR Accuracy	94% on degraded scans	1-2 years R&D
Historical Coverage	16 years (2008-2024)	Cannot be accelerated
Trust Network	NGO/journalist relationships	Years to build
First Mover	Operational platform	18-24 months

## SECTION 3: IMPLEMENTATION

### 3.1 Team

#### Chief Executive Officer

Strategy, fundraising, partnerships

- Technology and business development experience
- Regional market knowledge
- Investor relations

#### Chief Technology Officer

Technology architecture, AI/ML development

- Software engineering and machine learning
- Data engineering expertise
- Technical team leadership

#### Hiring Plan (with EIC Funding):

Role	Timing	Priority	Key Skills
ML Engineer (Corruption Detection)	Q1 2025	Critical	GNNs, anomaly detection
ML Engineer (NLP/OCR)	Q2 2025	Critical	Transformers, multilingual NLP
Full-Stack Developer	Q2 2025	High	React, Python, PostgreSQL
Data Engineer	Q3 2025	High	Scrapy, ETL pipelines
Business Development Manager	Q3 2025	High	B2B SaaS, regional experience

### 3.2 Work Plan and Milestones



Phase 1: Foundation (Months 1-6) - €500,000

Milestone	Deliverable	Success Criteria
M1.1	Launch paid subscription tiers	First 50 paid users
M1.2	Corruption detection MVP	3 algorithms operational
M1.3	Customer onboarding	100 paying customers
M1.4	Kosovo deployment	1,000+ tenders indexed

Phase 2: Regional Expansion (Months 7-12) - €700,000

Milestone	Deliverable	Success Criteria
M2.1	Albania platform launch	2,000+ tenders indexed
M2.2	Developer API launch	10 API customers
M2.3	Public transparency dashboard	Open access launched
M2.4	NGO partnership	MOU with TI chapter

Phase 3: Scale (Months 13-18) - €700,000

Milestone	Deliverable	Success Criteria
M3.1	Serbia platform launch	Largest market covered
M3.2	Montenegro coverage	5 countries total
M3.3	Enterprise sales	20 enterprise customers
M3.4	Advanced ML models	GNN collusion detection live

Phase 4: Growth (Months 19-24) - €600,000

Milestone	Deliverable	Success Criteria
-----------	-------------	------------------

M4.1	6th country launch	Bosnia and Herzegovina
M4.2	Revenue milestone	€2M ARR achieved
M4.3	User milestone	3,000+ paying customers
M4.4	Series A readiness	Investment materials ready

### 3.3 Risk Assessment

Risk	Probability	Impact	Mitigation
OCR accuracy plateau	Medium	High	Continuous training; human-in-the-loop
Slow customer adoption	Medium	High	Freemium model; partnerships
Government portal changes	High	Medium	Modular architecture; monitoring
Competition from larger players	Low	High	First-mover advantage; data moat
Political instability	Medium	Medium	Multi-country diversification

### 3.4 Financial Plan

#### Use of EIC Funds (€2,500,000):

Category	Amount	%	Key Activities
R&D - AI/ML	€1,000,000	40%	Corruption detection, GNN, OCR
Engineering	€600,000	24%	Platform, multi-country, API
Go-to-Market	€500,000	20%	Sales, marketing, partnerships

Operations	€250,000	10%	Legal, compliance, infrastructure
Contingency	€150,000	6%	Unforeseen challenges

Revenue Projections:

Year	Free Users	Paid Users	ARR
2025	5,000	200	€150K
2026	25,000	1,500	€900K
2027	60,000	4,000	€2.5M
2028	120,000	8,000	€5.5M

Path to Profitability: Q4 2026 | LTV:CAC Ratio: 8:1

3.5 Why EIC (Not Private Investment)

- 1. **Deep Tech Risk:** Corruption detection AI at scale is unproven. VCs require demonstrated unit economics.
- 2. **Long Time Horizon:** EU enlargement is a 3-5 year process. Standard VC exit timelines don't align.
- 3. **Public Good Component:** Free transparency features reduce commercial margins but maximize societal impact.
- 4. **Regional Focus:** Western Balkans is considered "frontier" by European VCs.
- 5. **Strategic EU Alignment:** EIC specifically funds EU enlargement-enabling technologies.

**Grant vs. Equity Preference:** Blended finance (grant for R&D, equity for go-to-market)

## Summary Statement

**The Problem:** €50+ billion in annual procurement across the Western Balkans is largely opaque - locked in scanned Cyrillic documents no technology can read. Corruption thrives in this darkness.

**Our Solution:** AI that reads what humans can't. We've built the first platform achieving 94% OCR accuracy on government scans, combined with semantic search and corruption detection algorithms.

**The Evidence:** 7,800+ tenders indexed, 450,000+ embeddings generated, operational platform processing data every 3 hours. TRL 7 - this is working technology, not a concept.

**The Ask:** €2.5M to scale from one country to six, advance corruption detection from TRL 6 to TRL 8, and build a sustainable business delivering lasting European impact.

**The Impact:** Every euro saved from corruption is a euro for schools, hospitals, roads. Every company winning on merit strengthens the economy. Every transparency tool brings the Western Balkans closer to Europe.

**We are not just building a company. We are building infrastructure for accountability.**

---

**NabavkiData d.o.o.e.l.**

Skopje, North Macedonia

*"Transparency through technology. Opportunity through data. Justice through intelligence."*

EIC Accelerator Step 1 Application