# From State-of-the-Art Methods to Practical Indexing Tools

## ArchXAI Workshop 3

24th of April 2026 • Riga, Latvia

Anssi Jääskeläinen | Tarvo Kärberg | Paul J. Aru

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI
XAMK
South-Eastern Finland University of Applied Sciences
Republic of Estonia National Archives
The National Archives of Finland
National Archives of Latvia

# Agenda

## Pre-Break

1. NER Module
2. PII Module
3. Tone Module
4. Demos

## Post-Break

1. Pilot Invite
2. Quiz
3. Deliverable
4. Q&A

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Why this matters for archives

**Find**

Key entities

**Detect**

Sensitive information

**Classify**

Document style or attitude

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Executive takeaway

| | |
|---|---|
| **Ready now** | Dedicated NER models |
| **Useful now** | PII detection and anonymisation |
| **Experimental** | Tone / sentiment classification |

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

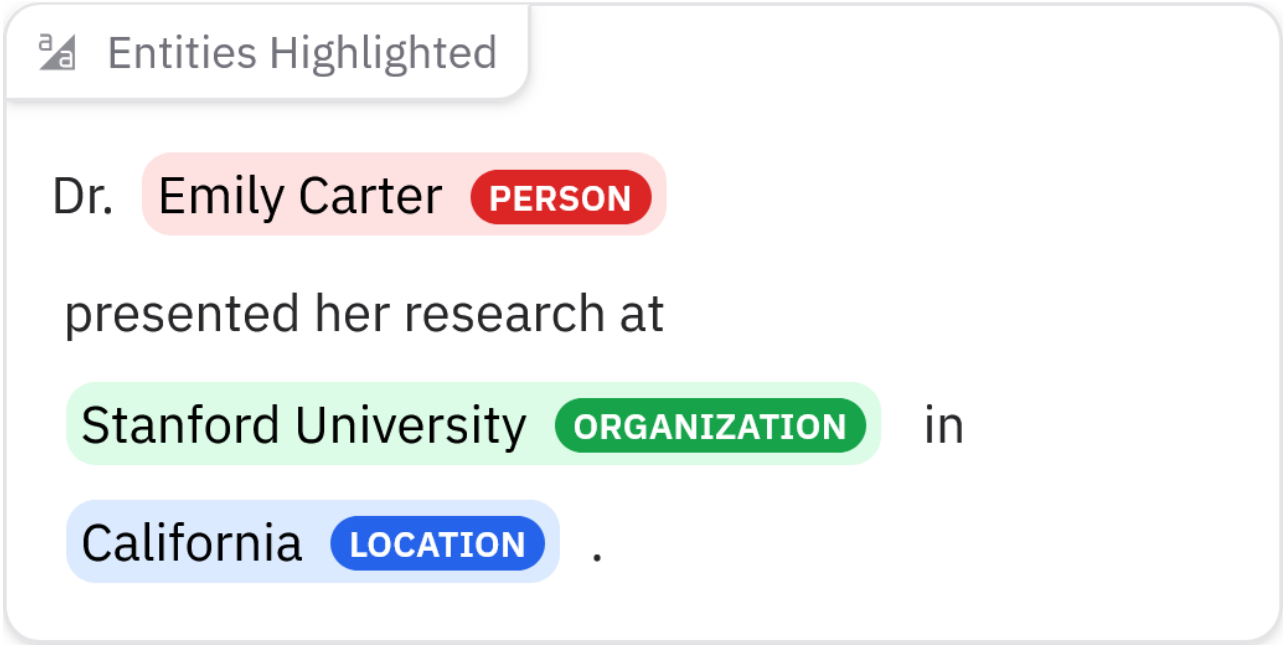
# Named Entity Recognition

## Part 1

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# NER: What & Why

- **Finds Key Entities in Text**
- **3 Core Types:**
  - PER
  - ORG
  - LOC
- **Text → metadata**
- **Improves search**




Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# NER: headline accuracy

| Language | F1 score across all entities |
|---|---|
| Estonian | 76% |
| Finnish | 75% |
| Latvian | 84% |
| Russian | 91% |

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# NER: Key Takeaways

## Dedicated NER models

- Default for bulk indexing
- Strongest balance
- Route by language

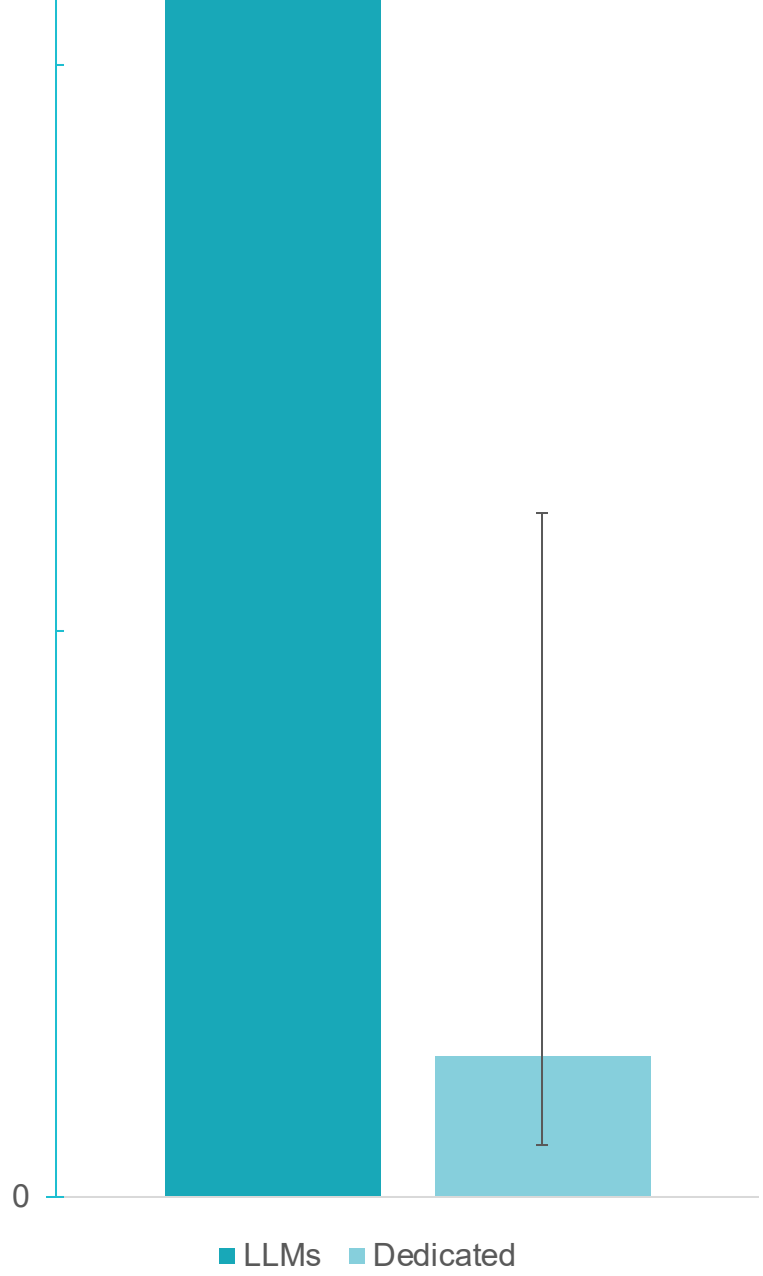
## LLM-based extraction

- Fallback for special cases
- Useful for new labels
- When speed isn’t relevant

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# NER: Speed at scale

- Dedicated: tens/sec
- LLMs: <1/sec
- LLMs are 100x-2000x slower




Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

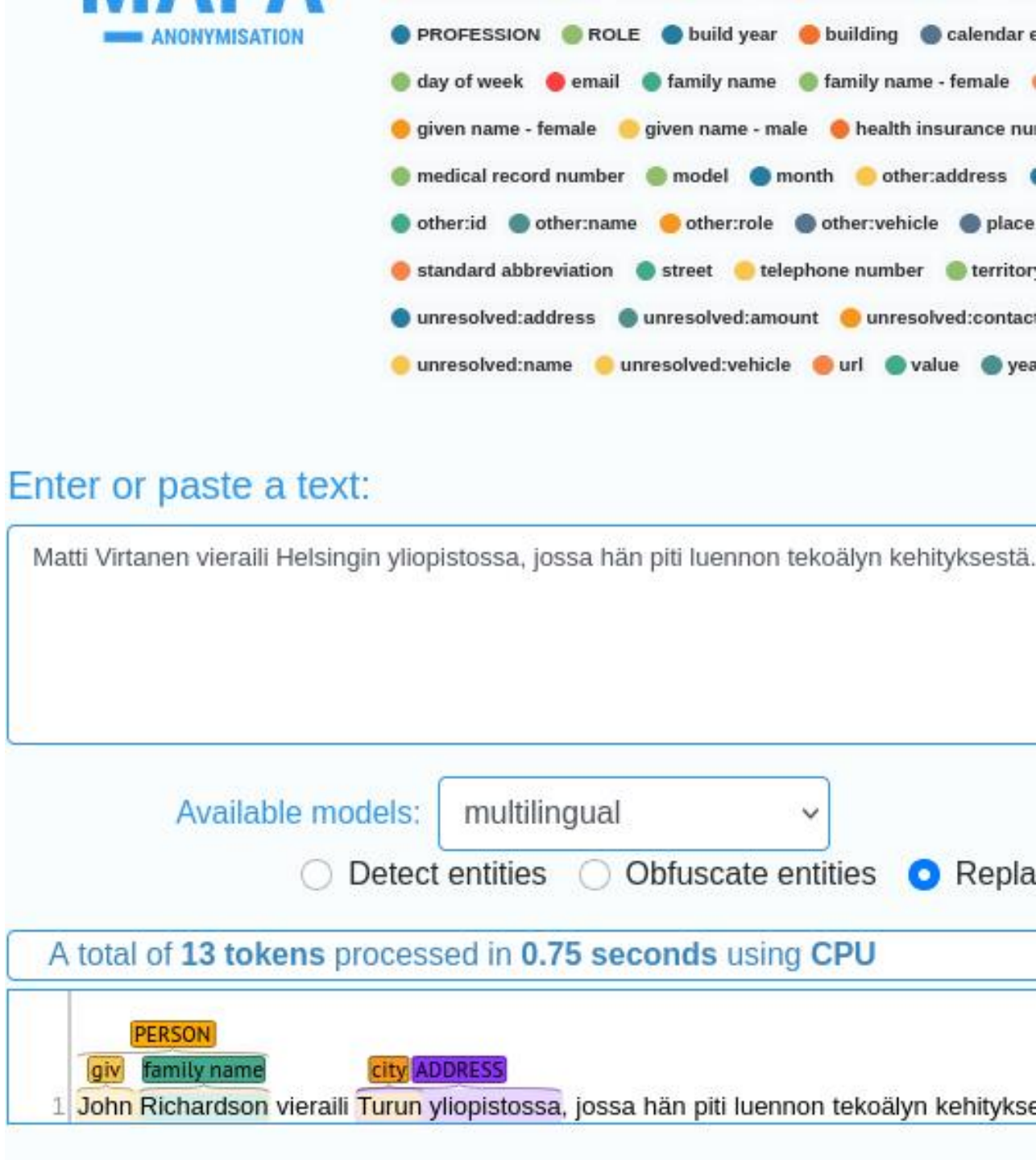
# Personally Identifiable Information

## Part 2

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# PII: More than NER

- **Goal: Legal Compliance**
- **Additional Types:**
  - Contact Details
  - Identifiers
  - Sensitive Spans
- **Access Decisions**





Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# PII: Recommendation

## Presidio - Default

For Building a Multilingual Archive Service or Backend Pipeline.

## MAPA - Visual Review

Where Anonymisation Behaviour and Visual Human Review Matter Most.

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Tone & Sentiment

## Part 3

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Tone: promising, but not ready

| Language | F1 score |
|---|---|
| Estonian | 34% |
| Finnish | 100% |
| Latvian | 60% |
| Russian | 69% |

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

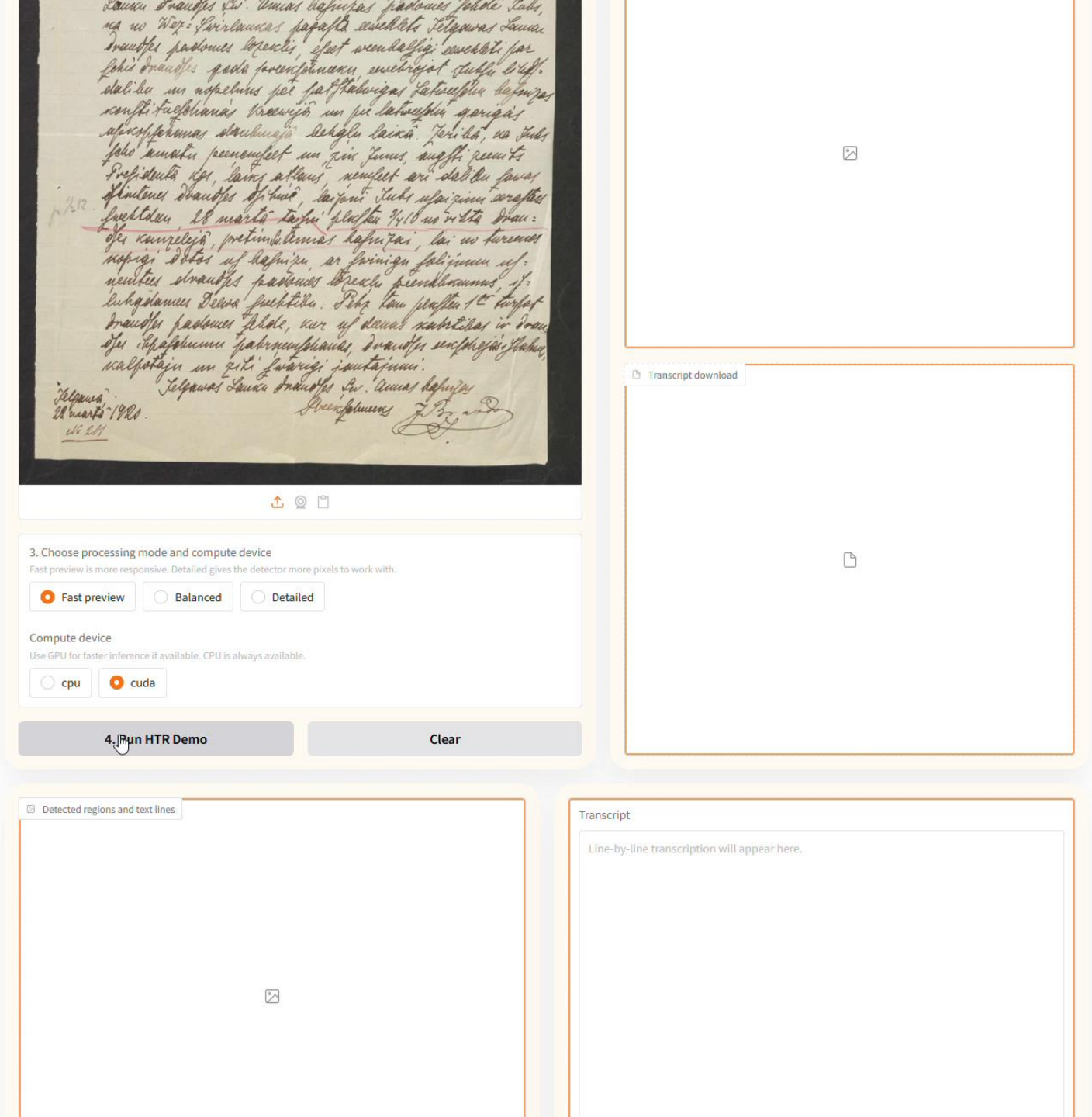
# Demos

## Part 4

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Demo

- Text Recognition
- First Indexing Modules
- Content Flow




Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Demo Links 🔗

**OCR**
**NER**

demot.memorylab.fi/archxai/latvia-demo

**HTR**

demot.memorylab.fi/archxai/latvia-demo-htr

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Coffee?

Interreg

Co-funded by the European Union

Central Baltic Programme

ArchXAI

# Pilot

## Part 5

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Join the Pilot

- **Main Principles**
  - Non-Disruptive Coexistence
  - Metadata Interoperability
  - Human-in-the-Loop Mandate
- **The 3-Phase Lifecycle**
  (Preparation, Execution, Evaluation)
- **How to Participate?**
  - Express your interest today
  - Help spread the word

tarvo.karberg@ra.ee

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

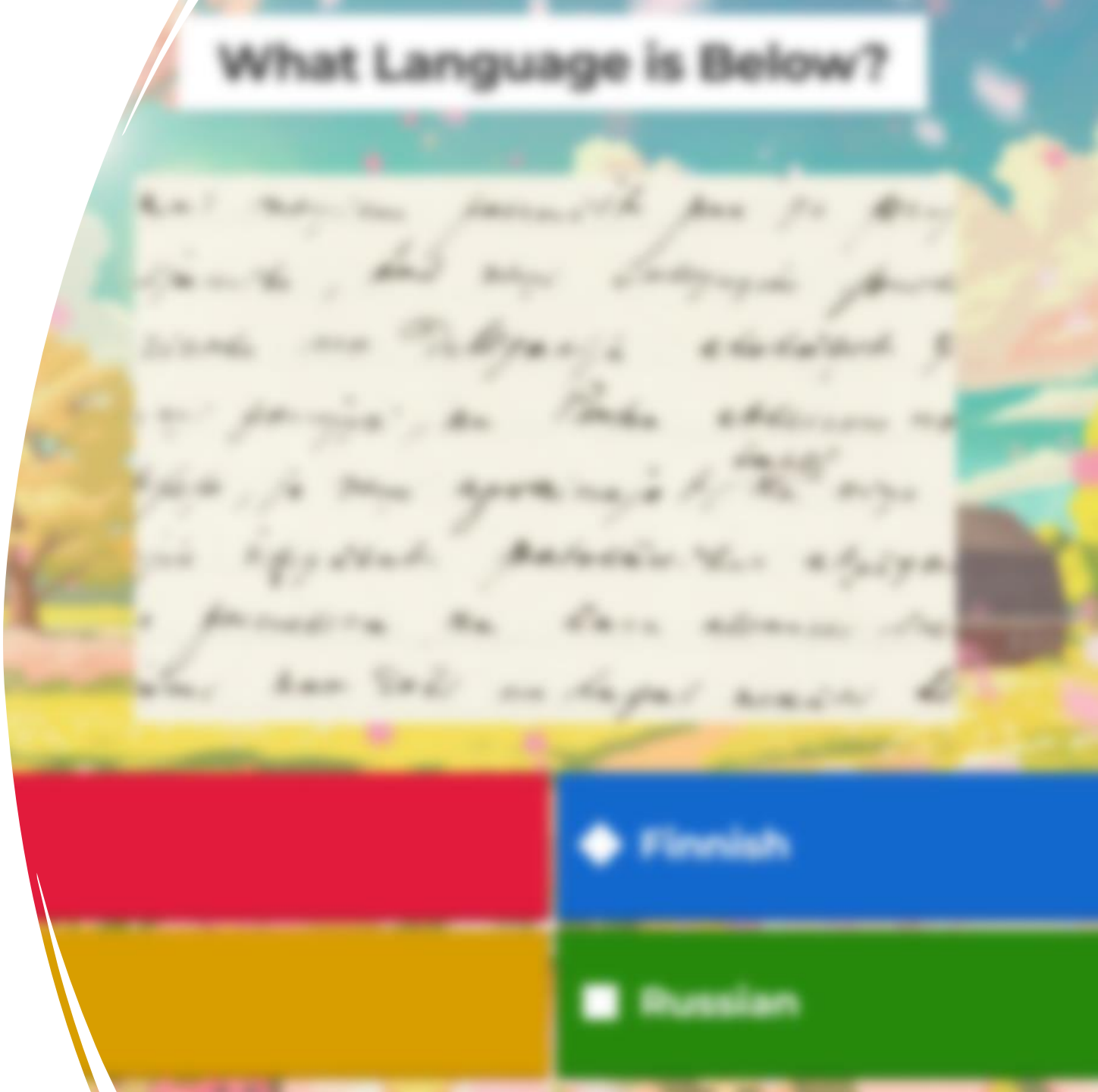
# Quiz

## Part 6

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Quiz

- Evaluate Our Score
- 1 team per table
- 1 min per question
- Reveal „*gold*“
  + note issues




Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Deliverable

## Part 7

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# D.2.1.1

- What’s included
- Where results live
- How updates happen

archxai.github.io

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Q&A

## Part 8

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# Q&A

- **What entities matter most?**
- **What's "*good enough*"?**
- **What data can we use?**

Interreg
Co-funded by the European Union
Central Baltic Programme
ArchXAI

# THANK YOU FOR YOUR PARTICIPATION

ArchXAI Workshop in Riga:
Evaluation Form

https://centralbaltic.eu/project/archxai/


Interreg
Co-funded by the European Union
Central Baltic Programme