Trustworthy enough for justice?

Evaluating risk assessment tools in European criminal justice systems

Authors: Victoria Burger and Merve Şahin

UNIVERSITÄT BAYREUTH

1. Are risk algorithms biased?

In the U.S., the **COMPAS system** gained notoriety for shaping bail and sentencing decisions through opaque, proprietary scoring — often with biased outcomes. Its internal workings are hidden, and independent studies have shown racial disparities in its predictions.

European tools are developed by public institutions and follow standardized scoring methods. They claim greater transparency and accountability, but often lack explainability and effective ways to contest results. Their trustworthiness cannot be assumed—it must be critically examined.

2. Comparison of the European systems

Tool	Focus	Main Purpose	Methodology	Strengths	Criticisms
STATIC-99R (Globally)	Individual	Whether a person will commit another sexual offense	Actuarial tool using 10 static personal variables	Empirically validated, widely used in criminal justice	Lacks dynamic risk factors; concerns about racial bias
HART (UK)	Individual	Whether a person will reoffend in general	Machine learning (random forest) using 34 variables	Incorporates broader social data; adaptive over time	Opaque "black-box" nature; potential for data bias
Precobs (Germany)	Location / Regional	Where a burglary is likely to happen	Crime pattern recognition (time, location, modus operandi)	Focused on property crimes; low false positive rate claimed	Profiling risk; effectiveness debated; limited scope

3. What makes a system trustworthy?

The idea of "trustworthy AI" has become central to European regulation, especially in the context of high-risk areas like criminal justice. But what exactly does trustworthiness mean?
Key elements include: *

- Transparency: Is the system understandable to those affected? Fairness and Bias Mitigation: Does the system prevent discriminatory outcomes? Does it reinforce existing inequalities?
- Accountability: Can decisions be traced, reviewed, and contested?
- Respect for Fundamental Rights: Are individuals' rights to due process and human dignity upheld?

* Trustworthiness criteria based on the EU AI Act, Articles 5 and 9. .

4. Evaluation of trustworthiness criteria

Criterion	Static-99R	HART	Precobs
Transparency	✅ Published rules	❌ Partial	❌ No public model
Explainability	⚠️ Moderate	❌ Low	❌ Very low
Contestability	⚠️ Rare in practice	❌ Limited	❌ Limited
Bias Monitoring	⚠️ Some validation,	❌ Not public	❌ Unknown
Human Oversight	✅ By trained professionals	✅ Claimed	✅ Claimed
Fundamental Rights Impact	⚠️ Subject to legal criticism	❌ Not clear	❌ Not assessed

Legend: ✅ Meets standard · ⚠️ Partially meets · ❌ Does not meet

5. Challenges

- **Bias and fairness**: How do data choices affect different social groups?
- **Explainability for whom?**: Are these systems understandable to subjects — or just to designers?

6. Conclusion

Not all European tools are trustworthy by default – as systems like Precobs and HART show. Legal framing helps, but real trust needs transparency, contestability, and accountability in practice.

Trust is not a label – it must be proven.