$ spacsec2026 --paper 23 // quantitative IS risk assessment

Application of Large Language Models to Information Security Risk Assessment via Simulation Modelling

Alexey Yu. Nazyvaev · Anatoly S. Minzov
National Research University “MPEI”
Department of Information Security Technologies, Moscow

CONFERENCE 4th SpacSec’2026
METHOD LLM → PERT → Monte-Carlo
CONTACT NazyvayevAY@mpei.ru

The problem: “high risk” has no price

ORDINAL SCORING · GOST R ISO/IEC 27005, FSTEC

+ simple, needs no data
– “high risk” has no price
– cannot justify a security budget

QUANTITATIVE · FAIR

+ answers in roubles
– needs loss statistics
– most organisations have too few incidents

THE SCALE HAS A CEILING

Top score ≈ **RUB 400 mln**.
NotPetya cost Maersk ≈
USD 300 mln.

RESEARCH QUESTION

Can a large language model replace the expert who sets the loss parameters?

The method

01
20 typical risks
phishing, ransomware, DDoS, insider, OT/CII…

→

02
Two LLMs, one prompt
Claude and Gemini act as risk analysts

→

03
Three numbers
min (5th percentile) · most likely · max (95th percentile)

→

04
PERT + Monte-Carlo
100 000 simulated years

→

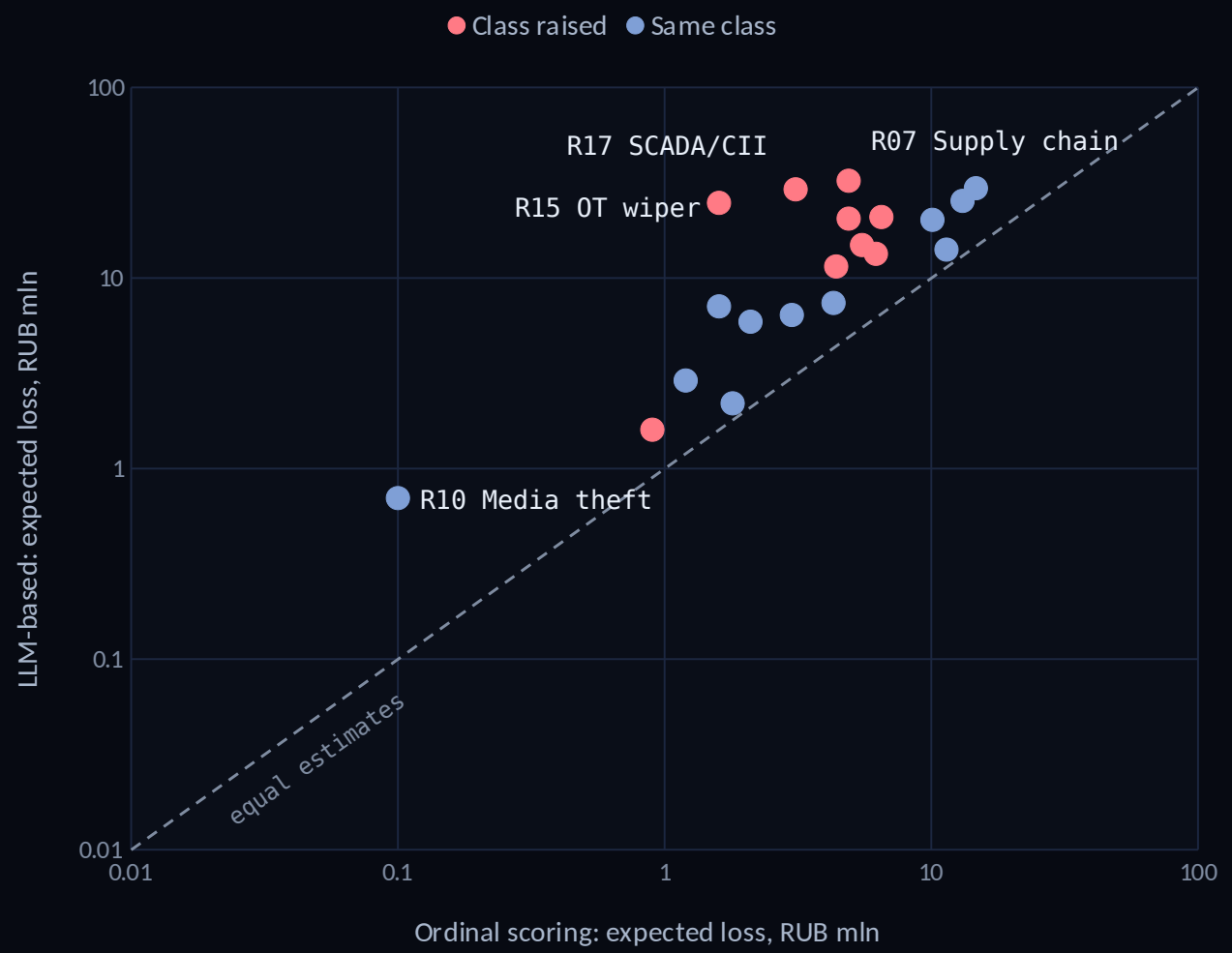
05
Compare
with ordinal scoring of the same 20 risks

PROMPT
7 loss components, from direct losses to long-term effects, plus a **real analogous incident**

EACH SIMULATED YEAR
did the incident happen? × how much did it cost?
Probabilities:
Verizon DBIR, **ENISA** 2024

MODELS DIFFER 5× OR MORE
7%
of cases, resolved by clarifying the query

Result 1: risk by risk



RANK CORRELATION

ρ = 0.70

moderate agreement

SAME RISK CLASS

11 / 20

other 9 went up, none down

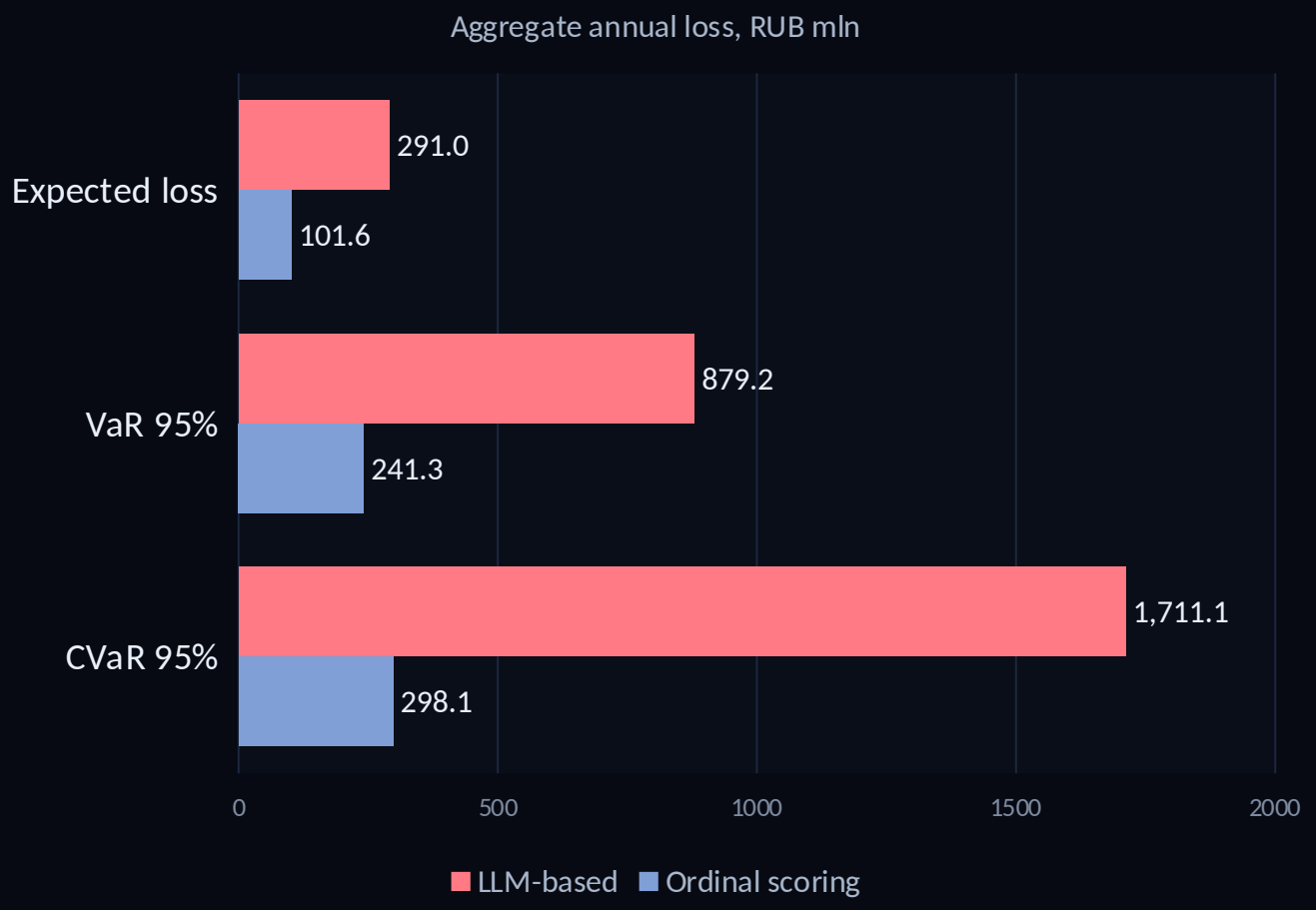
IN ROUBLES

The LLM estimate is **higher for all 20 risks**.

BIGGEST GAP · RARE, CATASTROPHIC

Destructive attack on OT: **×15.2**

Result 2: the whole portfolio



Expected loss	VaR 95%	CVaR 95%
×2.9	×3.6	×5.7

RARER AND HEAVIER → BIGGER GAP

VaR 95%: loss exceeded on average once in 20 years.
CVaR 95%: average loss in those worst years.
Insurance limits and reserves are set on these figures.

WHAT DRIVES THE RESULT

VaR 95% depends most on the LLM's **maximum loss** estimate. The number of simulations barely matters.

Limitations and takeaways

NO REAL LOSS DATA

We show **how far the methods diverge**, not which one is right.

SHARED KNOWLEDGE

Both models learned from similar public reports.

ERRORS DO NOT CANCEL OUT

A systematic bias adds up across risks.

TWO CHANGES AT ONCE

Parameter source and distribution changed together: we measure their combined effect.

TAKEAWAYS

- Loss estimate in roubles in minutes, not tens of expert hours
- All assumptions are explicit, auditable and cheap to change
- The result depends on the loss estimates, not on the simulation
- A research prototype, not a ready tool