



The Shell

Clinical Safety Briefing

A new standard for
Clinical AI safety



Dr Dom Pimenta

CEO,
Co-Founder TORTUS AI

Raising The Bar For Clinical Safety

At TORTUS, clinical safety is our number one priority.

This briefing outlines how we address clinical safety risks associated with generative AI, with a specific focus on hallucinations, and unpacks the proprietary system we've built to proactively mitigate those risks in day-to-day clinical use.

It is intended for clinical and digital leaders who need confidence that safety has been engineered deliberately, governed properly and proven in NHS settings.

We are proud to share this exciting new innovation with you.

Warm regards,

Dr Dom Pimenta
MBBS MRCP BSc (Hons)



The risk

Despite rapid advances in AI technology, hallucinations remain a known and unavoidable characteristic of large language models.

In healthcare, the risk is not simply that outputs may be incorrect, but that they may be fluent, plausible and presented without clear provenance. This makes errors harder to detect and increases the risk of inappropriate information influencing clinical judgement.

While clinicians will always retain responsibility for reviewing and approving clinical documentation, safe deployment of AI requires that they are supported by systems that minimise risk at source and provide a reliable base to work from.

Our Solution:



The Shell

The Shell is our proprietary, state-of-the-art hallucination removal system.

A safety control layer that sits as a protective 'shell' between AI generation and clinical review.

Its purpose is to ensure that only clinically appropriate, verified factual outputs are presented to clinicians.

In simple terms, The Shell is a double-checking agent that we've built to ensure that every single note generated by the TORTUS AI is checked back against the original transcription, proactively seeking out and correcting errors before they reach the clinician.

How The Shell works

The Shell operates through a three stage process:

1: Detect

Once an initial draft note has been generated from the audio transcription, it is assessed by The Shell to identify content that is unsupported by the source audio or falls outside agreed clinical parameters.

Any identified issues are flagged for correction before the note progresses further.

2: Edit

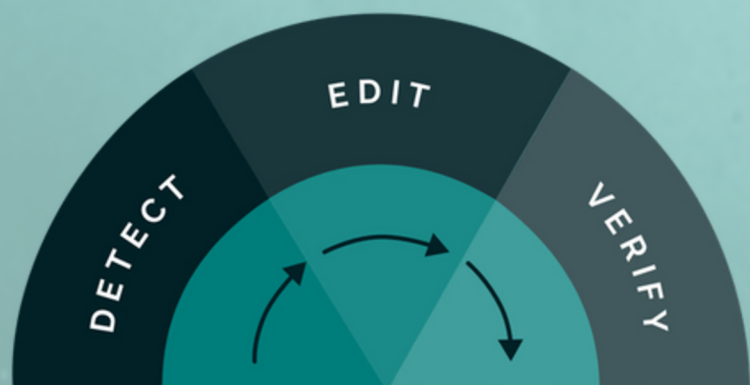
Where necessary, unsupported or hallucinated content is removed or corrected, and a revised draft is generated.

Accuracy metrics are recorded and outputs that fall outside predefined thresholds are reviewed by the TORTUS clinical team. This feedback loop supports continuous system improvement.

3: Verify

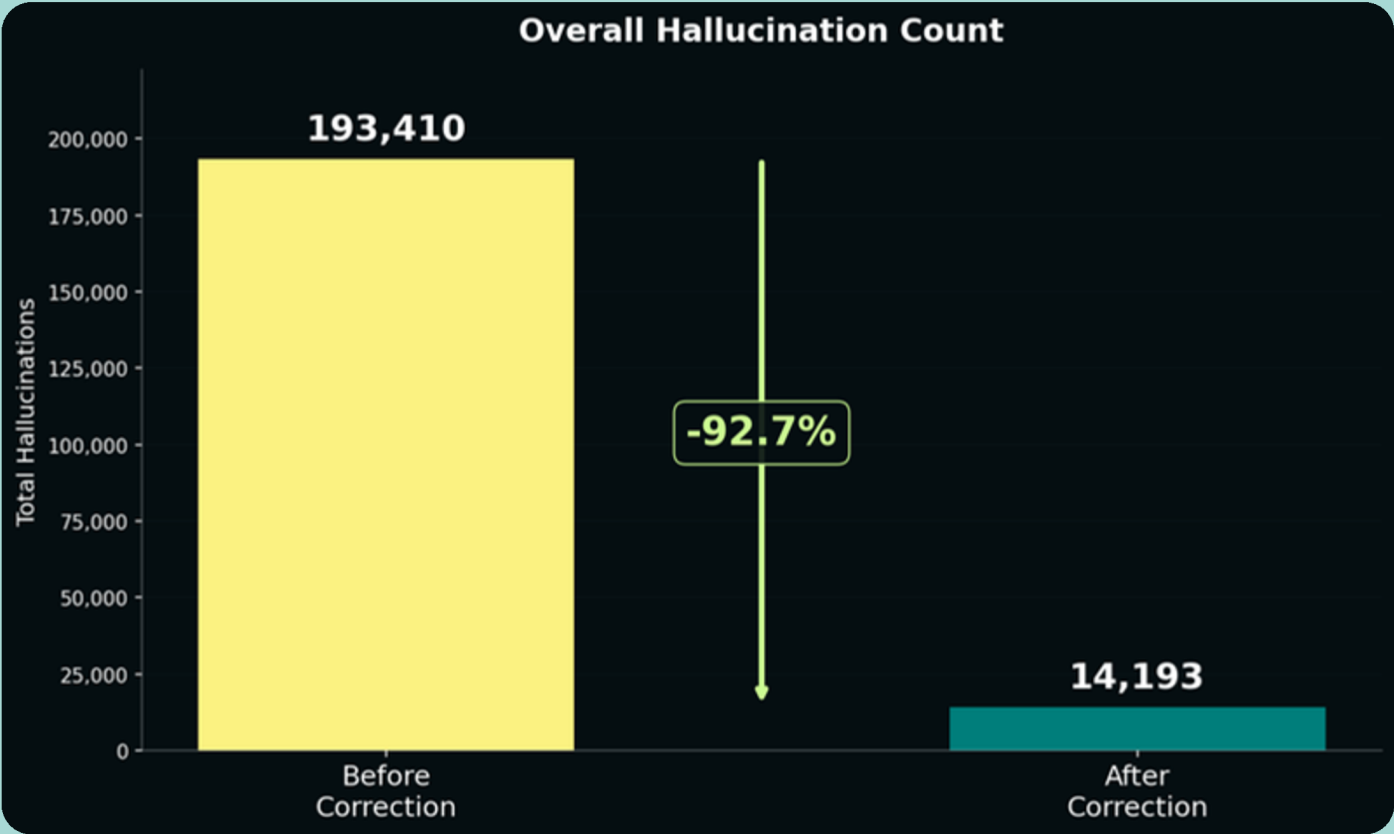
Before being surfaced to the clinician, the final draft is validated against the defined safety rules and source material. Clinicians then perform a final review and edit, retaining full clinical responsibility before the note is added to the patient record.

This entire process typically happens in as little as 2 seconds.



Impact in practice

The results speak for themselves - The Shell V1 is already able to spot and remove 92.7% of hallucinations, meaning even with this initial version TORTUS notes are almost completely **hallucination free**.

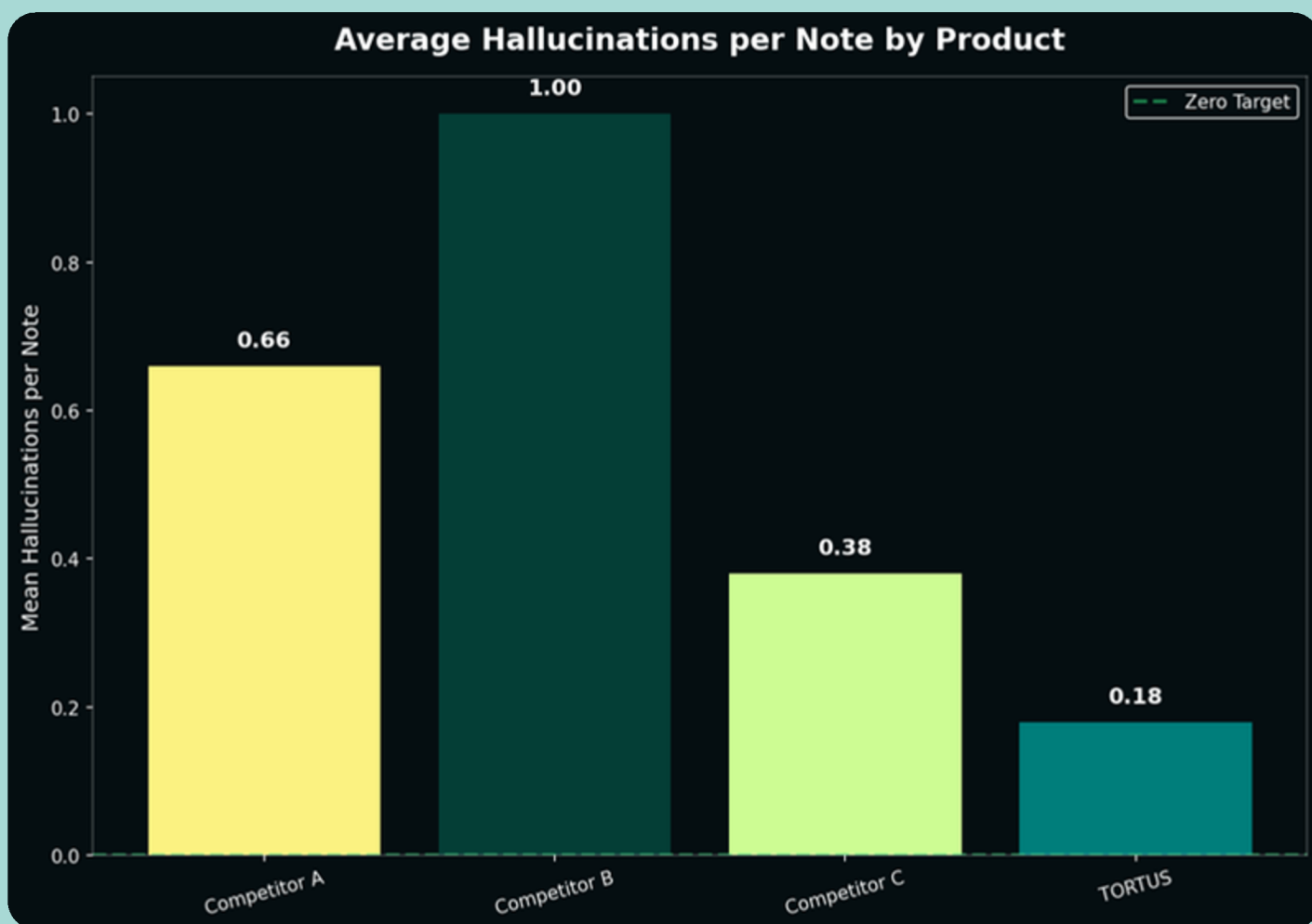


Hallucination counts before and after The Shell V1. Collected by applying The Shell to notes with detected hallucinations in production from Nov 2025 to Jan 2026 (over 150k).

These results are unprecedented in the market today, with the average vendor performance far behind (see below)

Product	Avg Note Accuracy	Std Dev	Min Note Accuracy	Halluc-Free %	Total Hallucinations	Hallucinations per Note
Competitor A	97.96%	0.035	86.70%	66%	33	0.66
Competitor B	97.09%	0.028	89.50%	36%	50	1.00
Competitor C	99.04%	0.019	91.90%	74%	19	0.38
TORTUS	99.58%	0.012	94.30%	86%	9	0.18

Summary Statistics for TORTUS’ SOAP note accuracy benchmarked against competitors across 50 consultations



In Jan 2026, we ran a comparative analysis against other clinical documentation AI systems. Using a controlled set of 50 general practice consultations, we generated notes in the standard SOAP format (Subjective, Objective, Assessment, Plan) across multiple platforms.

For clinicians, this means:

- Outputs are reliable by design, not by assumption.
- Trust is established through consistent system behaviour.
- Saving even more time by reducing edits required before finalising

For organisations, this means:

- Hallucinations are prevented upstream.
- Safety controls are auditable, observable and governable.
- AI use aligns with existing clinical safety expectations.

The difference

Competitor AI systems rely solely on “human in the loop” feedback, assuming clinicians will identify and correct all errors.

In practice, this shifts too much safety responsibility onto already stretched staff and increases risk.

The Shell takes a different approach.

Unsafe outputs are mostly prevented rather than just flagged. And hallucinations are treated as a system failure, not entirely a user responsibility.

Data safety

We take privacy and security incredibly seriously.

Unlike many other medical scribes, TORTUS never stores any patient data and all servers are run independently from the UK, and we don't train models on patient data.

Whilst we do monitor and measure metrics, these are only specifically to do with the accuracy of the system and never includes confidential information.



Summary

Generative AI can deliver real clinical value, but only if safety is treated as a first-order design requirement.

The Shell ensures AI assistance is dependable, verifiable and clinically appropriate before it reaches clinicians.

TORTUS systems are already deployed in live clinical environments, with safety operating as an active control rather than a promise. For further detail or a walkthrough of The Shell, please contact the TORTUS team.




Whether you're exploring
ambient AI, planning a pilot, or
shaping digital strategy, we'd
love to talk.



 Email

commercial@tortus.ai

 Check out our landing page

shell.tortus.ai

