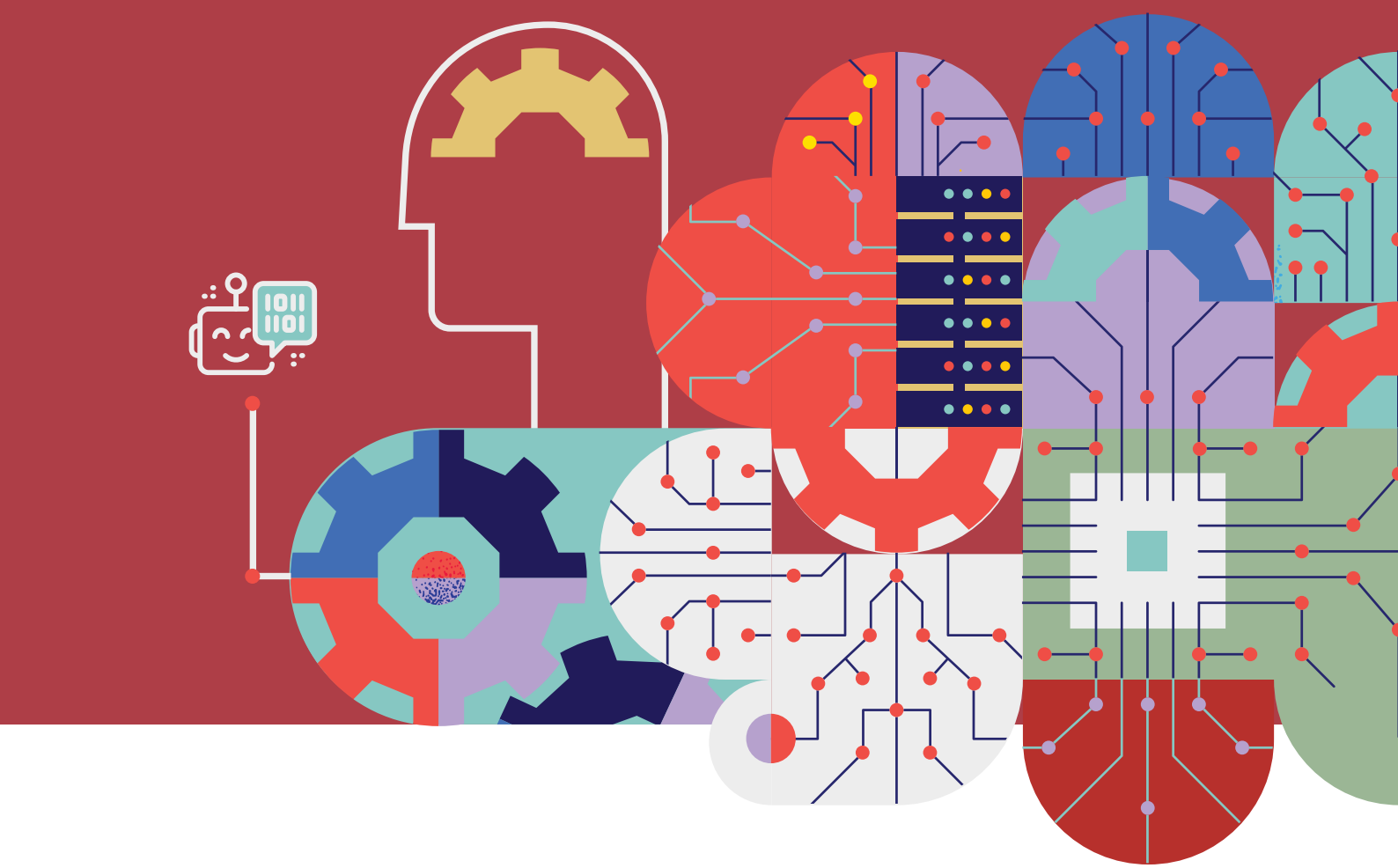
zensar



ZenShield (Guardrails to Avoid Jailbreaking)

Zensar's ZenShield (Guardrails to Avoid Jailbreaking) is a robust solution designed to enhance the reliability and security of gen AI applications by addressing challenges related to out-of-distribution queries, hallucinations, and jailbreaking. Zensar ensures that AI systems generate accurate, relevant, and safe outputs by implementing comprehensive guardrails and penetration testing.

Need for this offering

The rapid deployment of gen AI applications raises significant concerns about security and reliability, particularly regarding the potential for generating incorrect,

nonsensical, or malicious information. Traditional AI solutions may lack rigorous safeguards, compromising quality and user trust.

Features

- **Out-of-distribution query handling:** Detects and manages queries outside the model's training data
- **Hallucination prevention:** Mitigates instances where the AI generates false or nonsensical information
- **Jailbreak protection:** Implements robust content filters and penetration testing to prevent unauthorized manipulation
- **Real-time monitoring:** Provides continuous oversight to maintain the integrity of AI responses

Key differentiators

- Addresses potential vulnerabilities through proactive testing and safeguards
- Prioritizes security and prevents malicious or incorrect outputs more rigorously than competitors
- Tailors protection strategies for applications, ensuring robust security

Benefits

- Ensures more accurate and relevant AI responses
- Builds confidence in AI systems by preventing harmful or erroneous outputs
- Safeguards against manipulation and unauthorized access, protecting sensitive information
- Provides tailored security measures for diverse applications, from customer support to financial advisory

Use cases

- **Customer support chatbots:** Offers 24/7 support while preventing inappropriate responses
- **Personalized education tools:** Delivers customized content with safeguards against content manipulation
- **Healthcare advisory services:** Ensures secure handling of sensitive health information and advice
- **Content creation and summarization:** Prevents the generation of biased or false content
- **Legal and compliance automation:** Maintains the accuracy and integrity of legal documents and advice
- **Financial advisory services:** Complies with regulations and ethical standards in financial planning
- **Interactive entertainment and gaming:** Creates engaging content while preventing harmful or offensive material

zensar

An RPG Company

At Zensar, we're 'experience-led everything.' We are committed to conceptualizing, designing, engineering, marketing, and managing digital solutions and experiences for over 145 leading enterprises. Using our 3Es of experience, engineering, and engagement, we harness the power of technology, creativity, and insight to deliver impact.

Part of the $4.8 billion RPG Group, we are headquartered in Pune, India. Our 10,000+ employees work across 30+ locations worldwide, including Milpitas, Seattle, Princeton, Cape Town, London, Zurich, Singapore, and Mexico City.

For more information, please contact: **info@zensar.com | www.zensar.com**