

ORGANIZED BY:



WEBINAR

International AI Safety Report 2026



DATE

Monday,
May 25, 2026



TIME

03.30 – 05.00 PM
Jakarta Time



This meeting is being recorded. A closed captioning transcript is being generated.



OK

Leave meeting

View



ORGANIZED BY:



WEBINAR

International AI Safety Roundtable 2026



DATE

Monday,
May 25, 2026



TIME

03.30 - 05.00 P
Jakarta Time



Wing-Yee Lau (AISA)



KORIKA

KORIKA



Wing-Yee Lau (AISA)

Hammam Riza



Hammam Riza

Peserta_Febiola

Peserta_Febiola



BPH KORIKA

Remove spotlight

ORGANIZED BY:






WEBINAR

International AI Safety Report 2026

DATE
Monday, May 25, 2026

TIME
03.30 - 05.00 PM
Jakarta Time



Nezar Patria

This meeting is being recorded. A closed captioning transcript is being generated.



OK

Leave meeting



Peserta_Febiola

Peserta_Febiola



Aras Dewanto - USK



DAVID KARUNA GULDI



ORGANIZED BY:

AISII | **KORIKA** | **AISA** | **AISI** AI SECURITY INSTITUTE

WEBINAR

International AI Safety Report 2026

DATE Monday, May 25, 2026

TIME 03.30 - 05.00 PM Jakarta Time

Nezar Patria

ORGANIZED BY:

AISII | **KORIKA** | **AISA** | **AISI** AI SECURITY INSTITUTE

WEBINAR

International AI Safety Report 2026

DATE Monday, May 25, 2026

TIME 03.30 - 05.00 PM Jakarta Time

Wing-Yee Lau (AISA)

#AI TOWARD INDONESIA 2045

Webinar INTERNATIONAL AI SAFETY REPORT 2026

PROF. HAMMAM RIZA
PRESIDENT, KORIKA
PRINCIPAL ENGINEER, BRIN
PROFESSOR AI, USK

WEBINAR INTERNATIONAL AI SAFETY REPORT 2026

Exploring Global Insights,
Strengthening Indonesia's AI Safety Ecosystem

SPEAKERS

| | | | |
|--|--|---|---|
| <p>WELCOME REMARK</p>  <p>Hamмам Riza President of KORIKA</p> | <p>OPENING REMARKS</p>  <p>Nezar Patria Vice Minister of Communication and Digital Affairs</p> | <p>PRESENTATION ON THE INTERNATIONAL AI SAFETY REPORT 2026</p>  <p>Carina Prunkl Lead Author, International AI Safety Report 2026</p> | |
| <p>PANEL DISCUSSION</p> | | | <p>MODERATOR</p> |
|  <p>Aju Widya Sari Director of AI and New Technology Ecosystems, Komdigi</p> |  <p>Mateas Ramaputra Korika</p> |  <p>Carina Prunkl Lead Author, International AI Safety Report 2026</p> |  <p>Wing-Yee Lau COO & Product Director AI Safety Asia (AISA)</p> |

Organized by:






MONDAY, 25 MAY 2026 | 15.30 - 17.00 PM (JAKARTA TIME) WIB | VIA ZOOM MEETING

- Hamмам Riza
- Yolanda Febiola
- BPH KORIKA
- GKK
- Aras Dewanto - USK
- DAVID KARUNA GULDI
- Khairul Ikhsan



Indonesia must maintain the AI balance

AI is no longer only a research and innovation topic. It is already becoming strategic digital public infrastructure (DPI) that shapes services, productivity, sovereignty, and public trust.

“The question is not only what AI can do, but what kind of AI, Indonesia chooses to build.”

Innovation

Accelerate public services, healthcare, education, finance, logistics, and productivity.

Ethics

Prevent bias, disinformation, unequal benefit, and weak accountability.



Sovereignty

Align AI development with national interest, inclusion, and citizen dignity.

Participant video thumbnails for Hammam Riza.

How consumers view the transparency of AI interactions

As organizations progress to harness the benefits of AI, consumers, employees and citizens are watching closely and are ready to reward or punish behavior.

Clear, consistent, and understandable in its working
Transparent

Allows third-parties to asses data inputs and provide assurance that the outputs can be trusted
Auditable



Ability to explain the workings in language people can understand
Explainable and Interpretable



Ability to see how results can vary with changing inputs

Eliminates or reduces the impact of bias on certain users
Fair



62%

would place higher trust in a company whose AI interactions they perceived as ethical

61%

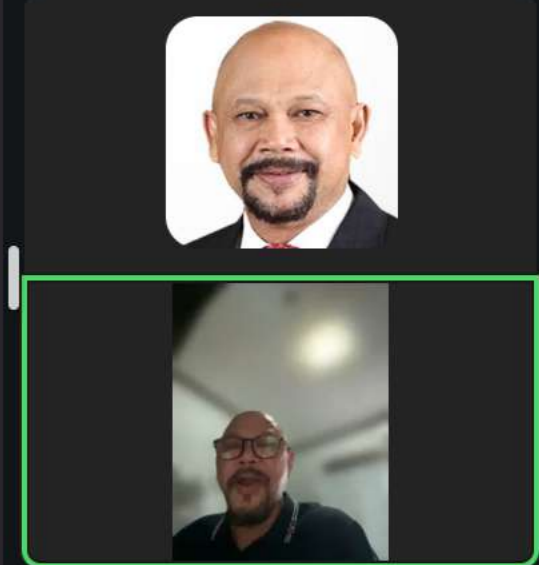
would share positive experiences with friends and family

59%

would have higher loyalty to the company

55%

would purchase more products and provide high ratings and positive feedback on social media.



AISII: THE SAFETY LAYER THAT MAKES INDONESIA'S AI FUTURE TRUSTED & SOVEREIGN

No sovereign AI without safe AI.



THE CHALLENGE

- AI is advancing rapidly across sectors
- But risks outpace governance
- Trust, safety, and sovereignty are at stake

AISII
AI SAFETY INSTITUTE INDONESIA
CROSS-CUTTING STANDARD & ASSURANCE

- Set Standards
- Evaluate & Assure
- Advise & Influence Policy
- Build Capacity & Collaborate

THE OPPORTUNITY

- Safe AI accelerates innovation
- Strengthens digital sovereignty
- Positions Indonesia as a leader in responsible AI for the world

AISII STANDARDS, TOOLS & ASSURANCE APPLIED ACROSS ALL KORIKA PROGRAMS

VP TALENT DEVELOPMENT & ACADEMY
Building Indonesia's AI Talent Engine

- KORIKA ACADEMY**
Education & Certification
- TALENT PIPELINE PROGRAM**
Develop & Grow AI Talent
- AI CERTIFICATION BODY**
Competency & Safety Aligned

VP INNOVATION & IMPACT
Driving Solutions & Public Value

- AI RESEARCH & INNOVATION LAB**
Research & Local AI Development
- KORIKA IMPACT**
AI for Public Good & National Priorities
- AI SANDBOX & TESTBED**
Test, Validate, D-risik

VP BUSINESS & INDUSTRY
Enabling Adoption, Scaling & Economic Impact

- KORIKA CONSULTING**
Strategy & AI Transformation
- INDUSTRY PARTNERSHIP OFFICE**
Partnerships & Ecosystem
- AI VENTURE & COMMERCIALIZATION HUB**
Scale & Create Economic Value

OUTCOMES FOR INDONESIA

- TRUSTED AI ECOSYSTEM**
All AI systems are safe, reliable, and accountable
- DIGITAL SOVEREIGNTY**
Indonesia's data, models, and values are protected
- GLOBAL COMPETITIVENESS**
Innovation with confidence and global credibility
- LEADERSHIP IN AI SAFETY**
Indonesia as a reference for the Global South

AISII OKR ROADMAP 2026

| | | | |
|--|---|--|---|
| <p>Q1 JAN - MAR FOUNDATION & ALIGNMENT</p> <ul style="list-style-type: none"> Establish AISII governance & operating model Approve National AI Safety Framework v1.0 Define risk taxonomy & assessment methodology 2 pilot assessments initiated | <p>Q2 APR - JUN STANDARDIZATION & PILOTING</p> <ul style="list-style-type: none"> Release AI Safety Standard v1.0 Launch evaluation toolkit & red-teaming protocol 5+ pilots across all VPs assessed Safety training for 200+ professionals | <p>Q3 JUL - SEP INTEGRATION & SCALING</p> <ul style="list-style-type: none"> Integrate AISII standards into all VP programs 20+ systems/models evaluated AI Safety Certification (Level 1) launched Policy recommendations submitted | <p>Q4 OCT - DEC MATURITY & GLOBAL POSITIONING</p> <ul style="list-style-type: none"> Release AI Safety Standard v2.0 50+ systems/models evaluated 500+ professionals certified International recognition & partnerships |
|--|---|--|---|

AISII ENSURES THAT EVERY STEP OF INNOVATION IN INDONESIA IS SAFE, TRUSTED, AND ALIGNED WITH OUR NATIONAL VALUES.

8

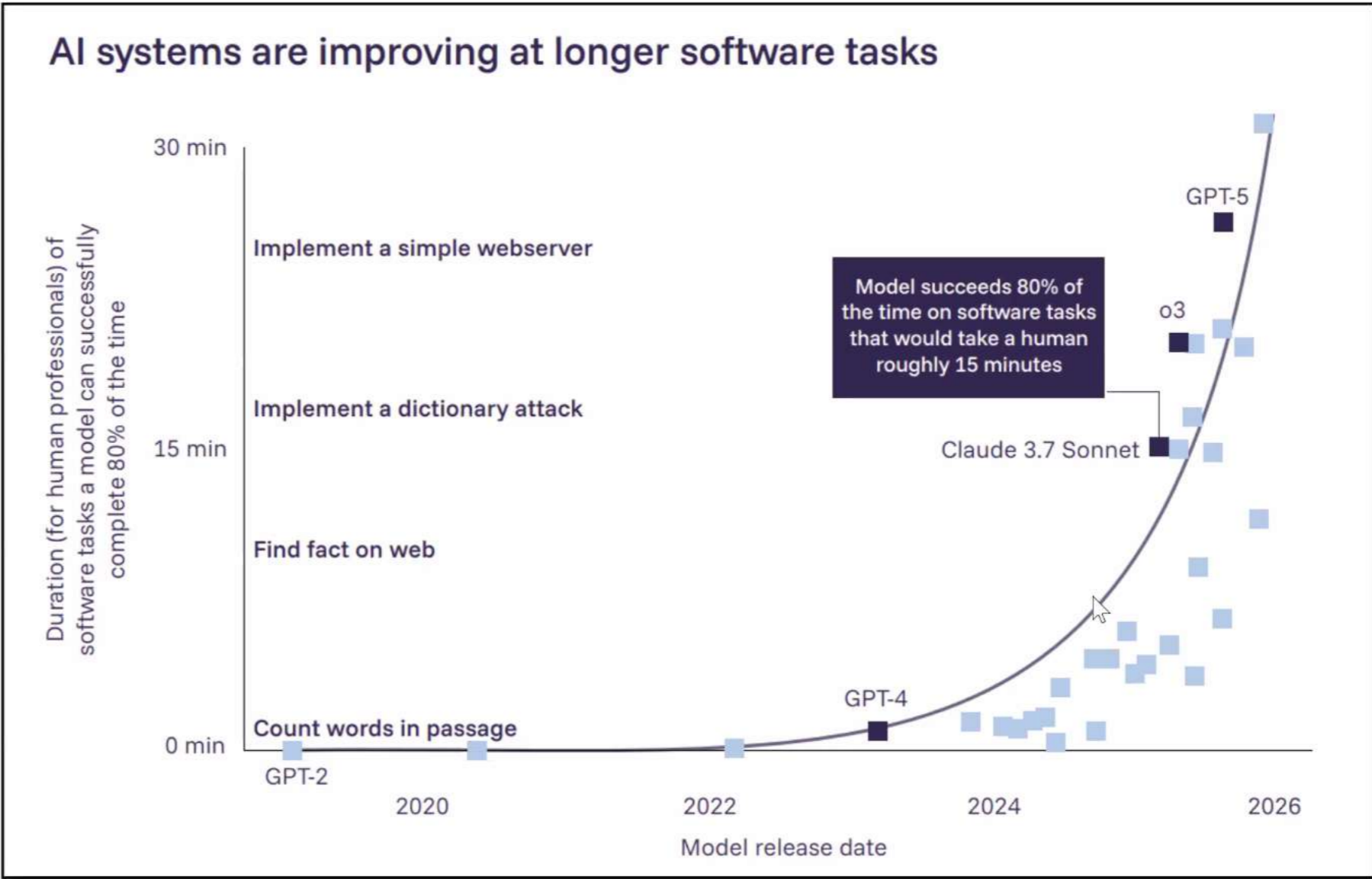
Longer, autonomous tasks now possible by agents

1

Measured **task length** has roughly tripled.*

2

Progress expected to continue



*) Measured at ~80% success on software engineering tasks of known human duration (METR)

Figure source: Kwa et al., 2025

Remove spotlight

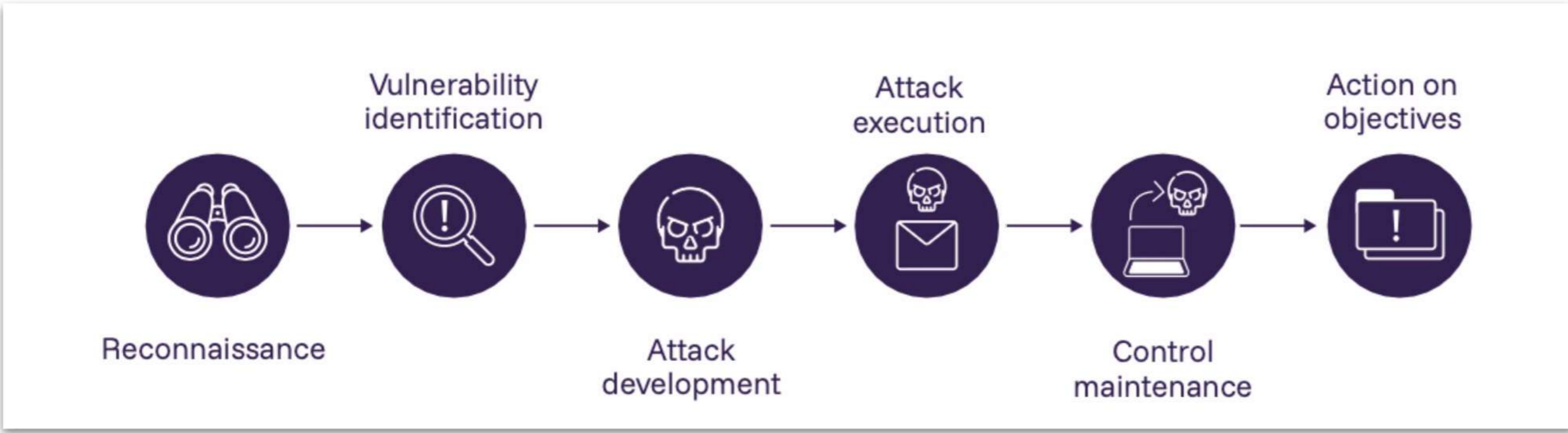
Carina Prunkl

Capabilities are difficult to measure

- Evaluation results depend on elicitation: prompting, scaffolding, tool access, time, and inference compute
Examples: third parties used scaffolding techniques to elicit greater cyber capabilities than were measured in pre-deployment testing
- Benchmarks can saturate, become contaminated, or reward benchmark-specific optimisation
Examples: Claude Opus 4.6
- Governance implication: a model that appears below a risk threshold in one setup may exceed it in another



AI is increasingly useful for cybercrime



Better at **detecting software vulnerabilities**

AI agents identified 77% of flaws in major cybersecurity competition

Malicious actors now **regularly use AI**

All major developers now report on criminal use

Figure source: Adapted from Rodriguez et al., 2025

Remove spotlight

Carina Prunkl

But challenges remain...

These challenges create an **'evidence dilemma'** where the risk landscape changes rapidly but the evidence about new risks and mitigation strategies is slow to emerge

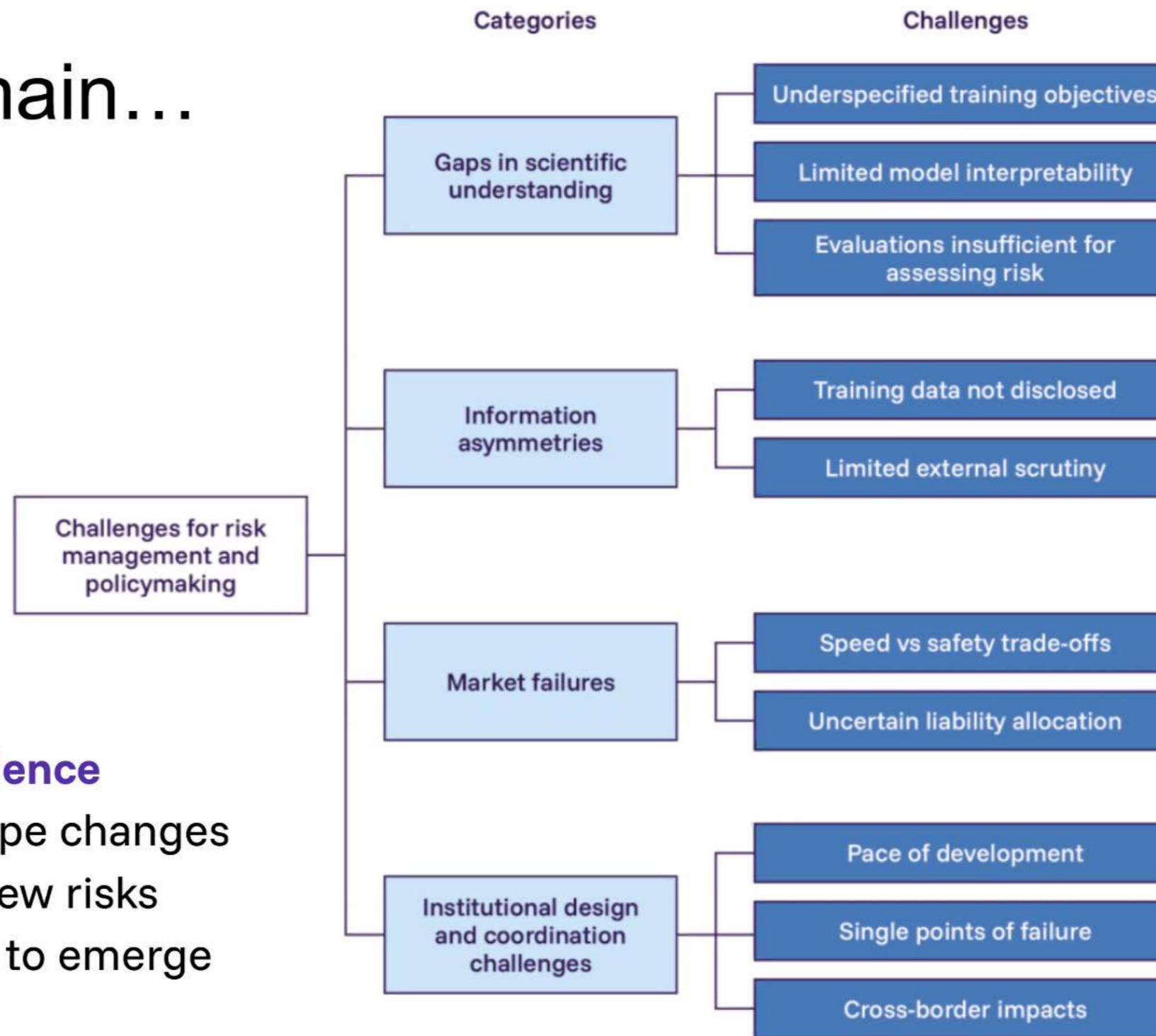


Figure source: Zou et al., 2025, cited in Anthropic, 2025



Safeguards remain limited in practice

- Individual safeguards have improved but **remain brittle**
- Attackers can still **frequently bypass safeguards**

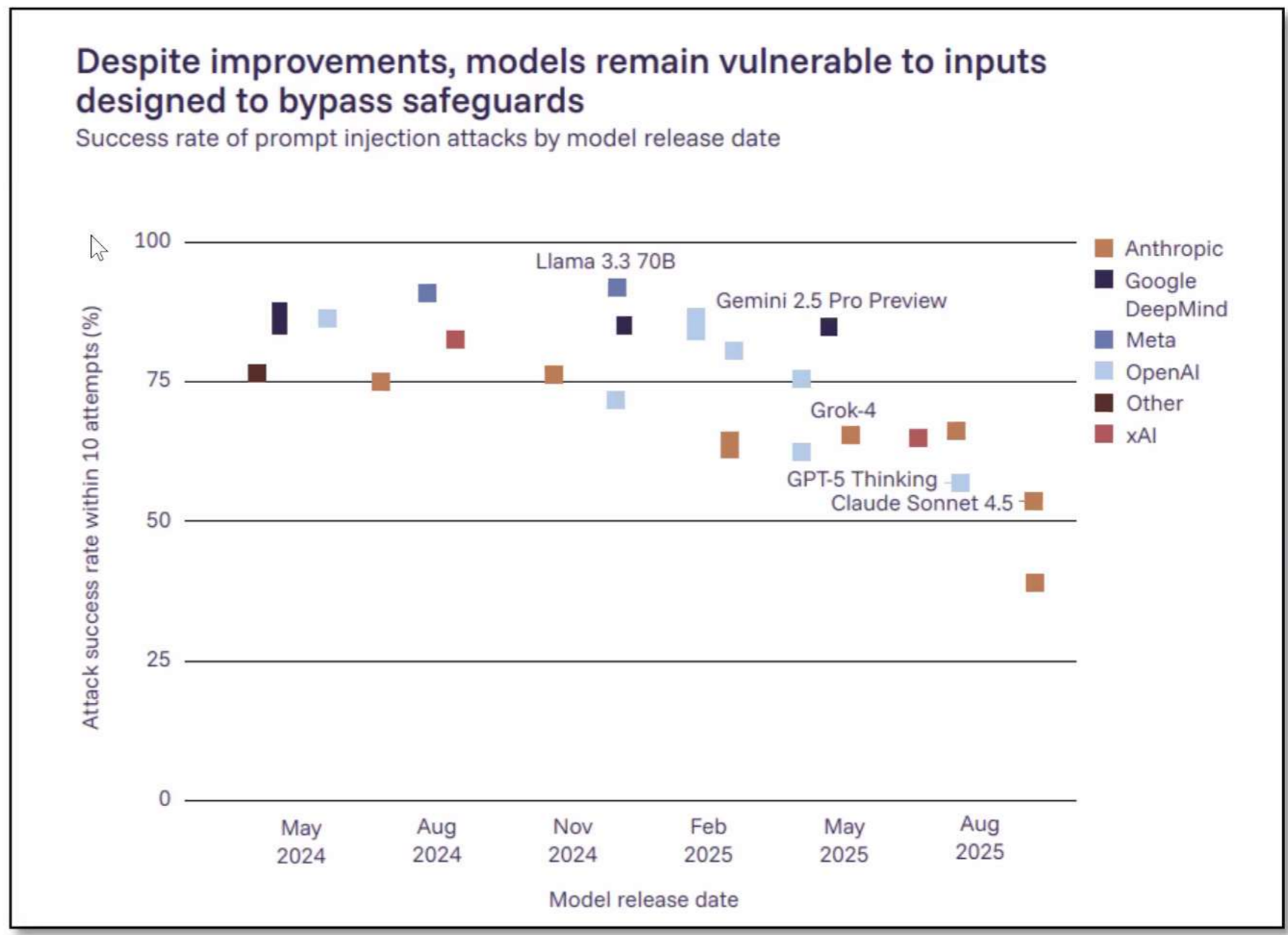


Figure source: Zou et al., 2025, cited in Anthropic, 2025

Remove spotlight

Carina Prunkl

This meeting is being recorded. A closed captioning transcript is being generated.



OK

Leave meeting



| | | | | | |
|---|-------------------------------|--------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | KORIKA | | | | |
| Hammam Riza | KORIKA | BPH KORIKA | DAVID KARUNA GULDI | Dian Andriana | Ask to unmute |
| | | | | | |
| Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 |
| Ali_USK | SYAUKAS RAHMATILLAH | Asril | Rahmat Wibowo_InfraLoka | Yolanda Febiola | |
| | | | | | |
| Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 |
| Sugeng Miftahudin | Fanny Anggraeni | Samuel - Synnex Metrodata I... | M Miqdad Ramadhan F | A. Djalal Latief | |
| | | | | | |
| Tri Sundara Webinar AI Safety Report 2026 | | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 | Webinar AI Safety Report 2026 |
| Sugeng... | | Hariyono Darmawan | Sudiono | Ridwan Suhud | |
| Bayu Nur Angga | Herman Agustiawan | | | | |

Liene Lia... 1/6

andre

thumbs up icons

1/6

thumbs up icon

| | | | | |
|--------------------------|--------------------|---------------------|---------------|-----------------|
| Fawad Karimi_U... | Tri Sundara | Randy - AISA | | Bellinda |
| Fawad Karimi_USK | Tri Sundara | Randy - AISA | Fahmi Adestya | Bellinda |

| | | | | |
|--|--|------------------------------------|-----------------------------------|---|
| <p>International AI Safety Report 2026</p> <p>DATE: Monday, May 25, 2026 TIME: 03.30 - 05.00 PM Jakarta Time</p> <p>BPH KORIKA</p> | <p>Yolanda Febiola</p> | <p>GKK</p> | <p>Aras Dewanto - USK</p> | <p>DAVID KARUNA GULDI</p> |
| <p>Carina Prunkl</p> | <p>Webinar AI Safety Report 2026</p> <p>DATE: Monday, May 25, 2026 TIME: 03.30 - 05.00 PM Jakarta Time</p> <p>Khairul Ikhsan</p> | <p>Nezar Patria</p> | <p>Wing-Yee Lau (AISA)</p> | <p>Danial Alfayyadh</p> <p>Ask to unmute</p> |
| <p>Webinar AI Safety Report 2026</p> <p>DATE: Monday, May 25, 2026 TIME: 03.30 - 05.00 PM Jakarta Time</p> <p>Ali_USK</p> | <p>M. Dirhamsyah</p> | <p>Annifa Iqramitha</p> | <p>Hammam Riza</p> | <p>SYAUKAS RAHMATILLAH</p> |
| <p>Webinar AI Safety Report 2026</p> <p>DATE: Monday, May 25, 2026 TIME: 03.30 - 05.00 PM Jakarta Time</p> <p>Dian Andriana</p> | <p>Leontinus - Kemenko PM</p> | <p>BERTHAN RIDHO RAMA</p> | <p>Ramsha Khan</p> | <p>M.Yani-Med Fac-USK</p> |
| <p>Rafael Ibrahim - AJARI.AI</p> | <p>Hariyono Darmawan</p> | <p>Dr Irwansyah Tanjung</p> | <p>Sherly Haristya</p> | <p>Peserta RPS</p> |

This meeting is being recorded. A closed captioning transcript is being generated.



OK

Leave meeting

View

Webinar AI Safety Report 2026

Ali_USK

International AI Safety Report 2026

BPH KORIKA

Webinar AI Safety Report 2026

SYAUKAS RAHMATILLAH

Webinar AI Safety Report 2026

Hariyono Darmawan

International AI Safety Report 2026

Yolanda Febiola

Webinar AI Safety Report 2026

M Miqdad Ramadhan F

International AI Safety Report 2026

A. Djalal Latief

Webinar AI Safety Report 2026

Sudiono

Hammam Riza

Hammam Riza

Fanny Anggraeni

Fanny Anggraeni

Herman Agustiawan

Rinaldi Udra Suhanda

Adelia Shintia Ningrum

Anis F

Anis F

M. Dirhamsyah

BERTHAN RIDHO RAMA

RMCC

ftriantoro

Adi Imam

Rizal Bachtiar

Rizal Bachtiar

Nafriza Hafid R...

Nafriza Hafid Rahadian R

Iwan S

Rahmat Darmaw...












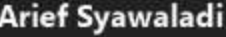












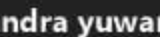
Rahmat Darmawan

Fachrizal -

This meeting is being recorded. A closed captioning transcript is being generated.

OK Leave meeting

View

| | | | | |
|--|---|--|---|--|
|  Dinda Amelia |  Peserta RPS |  Peserta_Febiola |  dwie |  Peserta |
|  Rahmat |  Fenny Desriani |  Al Muchlisin |  Umar Zukhran |  Eylien - WEGE |
|  Arief |  Arief Syawaladi |  windi siti maryam |  Chandra Yedija K |  Atjih Sukaesih |
|  Azka Tsabbita |  Triyakom - Windy |  ALEXANDER B |  Satria Wibawa |  Mega S |
|  K. TANJUNG - WEGE |  Akbar F |  M. Reza Alfin |  A |  sandra yuwana |

2/3

2/3

Unmute Stop Video Security Participants 74 Chat 34 Share Screen Stop Summary AI Companion Record Show Captions Breakout Rooms Reactions Apps Whiteboards Leave

KORIKA



Paramita Putri Wijayanti



Abdulla Emir Pramudya :)

boby harmoko



Fadlan Nuril Fahmi_Humas Komdigi

dwie

Harun Faizal



Peserta

Surya

Yadi

dwie

Harun Faizal

Surya

Yadi



Arief Kusuma



Dr Irwansyah Tanjung

Farhan R

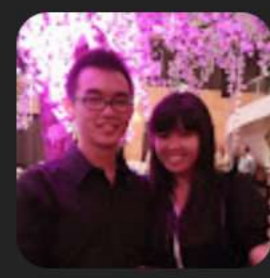
Farhan R

M. Fahriadi

M. Fahriadi

Oskar Riandi

Oskar Riandi



Boby Hartanto

Whidia_KemenkoPM

Whidia_KemenkoPM



Peserta RPS

Fawad Karimi

Fawad Karimi



Aldy - USK



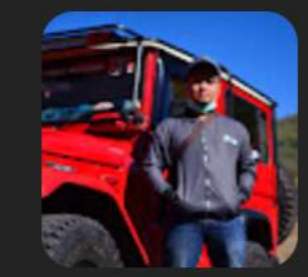
Ivan Alexander

nafisa

nafisa



Emiria Dinar Triana



putut nalendra



M.Yani-Med Fac-USK



M.Yani-Med Fac-USK



Muhammad Khalid Al Ghifari



Liene Lianawati



Jodi

eni ika

eni ika



Cut Sula Fhatia Rahma



Siti Rusdiana

Nur Fitriana

Nur Fitriana

Shaski

Shaski

Maha Rani

Maha Rani



Wahyu Cahyadi -- ワヒュ-チャヤディ

Derry Allen K.

Derry Allen K.



Anto S Nugroho



Rahmat

Leontinus - Kemenko PM

Leontinus - Kemenko PM



Doni F



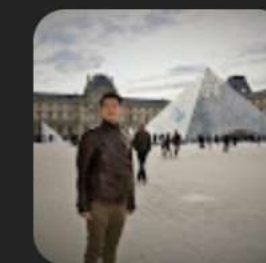
Sherly Haristya



Gandhi KH



Cut Jihan Shavira



Heng Lie

Christian

Christian

Matheace Ramaputra

Matheace Ramaputra



Jonathan Gonzalez



Arifka Yusri



Suci AIKITA

Mochamad Azhar

Adam Pantouw

Agung Tri Nugraha

Fenny Desriani



Mochamad Azhar

Adam Pantouw

Agung Tri Nugraha

Fenny Desriani

Aju Widya Sari Komdigi

Rizki Triyani



Agoeng Srimoeljanto



Rizki Triyani

Savitri Amalia_AIKITA-Komdigi

Fahmi Adestya

Agoeng Srimoeljanto

solfarina



5/8



Indah Sari

Indri maria



Sri Agustiningtyas

M. Arief - BRIN

Indah Sari

Indri maria

Dwi Ahmad_USK



5/8

Carina Prunkl



Ann M

Ramsha Khan

Carina Prunkl

Saiful Ridwan

Edward Tsoi

Ann M

Ramsha Khan

Anonyo Mitra



EKRAF_Dandy Yudha

Anonyo Mitra

Erwin

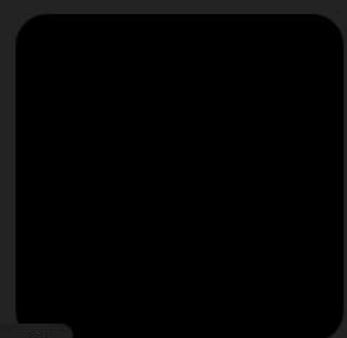
Luthfiyyah Tiurrana Putri

Steven Setiawan

EKRAF_Dandy Yudha



Ali Reza Syariati



Dinda Amelia



Iwan S

Ananto

Yunarso Anang



Fachrizal -



Andreas Tjendra

Fathur



shafa



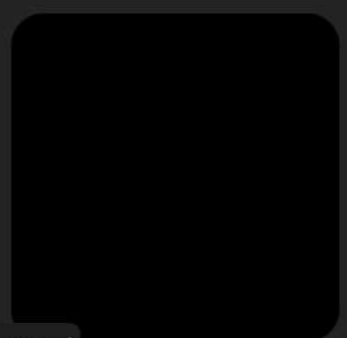
Hendar



Egatama Arista



Anzo



AFTECH_Suci

Rahmat Darmawan

Yosi Sahreza



Al Muchlisin

Rafli Albhukhori



Heru Setyabudi

Umar Zukhran

Gembong sw

Gembong sw

Umar Zukhran

Randy - AISA

Andri Puji Prasetiyo



Naja Safira



Bramajaya



Eylien - WEGE