Model G20 Summit 2027

STUDY GUIDE

Hosted at FLAME University, Pune · 18 to 20 June 2027

Arrive by the evening of Thursday 17 June, leave Monday 21 June



Sherpa track

Digital Economy and Artificial Intelligence

Should G20 members agree common safety tests for the most powerful AI models before they are released?

AT A GLANCE

US$4.8 trillion

The size UNCTAD projects for the global AI market by 2033, up from US$189 billion in 2023[1]

40%

Share of the world's AI research and development done by just 100 companies in 2022[1]

10^{25}

Computing operations used in training above which the EU presumes an AI model has high-impact capabilities[2]

YOUR GROUP MUST DECIDE

1. Which models count as the most powerful
2. What a test must check, and who runs it
3. Whether testing comes before release, after it, or both
4. Whether testing is voluntary or required by law
5. Who reports progress, and by when

FIVE TERMS TO KNOW

Frontier AI
Highly capable general-purpose AI models at the leading edge[3].

Pre-deployment testing
Testing an AI model before it is released to the public.

AI safety institute
A government body that tests leading AI models and studies their risks[4,5].

Systemic risk
In EU law, the risk from a model with high-impact capabilities[2].

Adversarial testing
Trying on purpose to make a model misbehave[6].

INSIDE THIS GUIDE

1	Letter from the chairs	1	2	How the G20 and this simulation work	1	3	The issue explained	2
4	What has happened so far	4	5	Facts that matter	5	6	Core tensions	5
7	Case studies	6	8	Country positions	7	9	Options on the table	10
10	What the communiqué must answer	11	11	Preparation	12	12	Glossary	13

1. UNCTAD. "AI market projected to hit $4.8 trillion by 2033, emerging as dominant frontier technology" (Technology and Innovation Report 2025). 2. European Union. Regulation (EU) 2024/1689 (Artificial Intelligence Act), Official Journal L, 2024, EUR-Lex. 3. GOV.UK. "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023". 4. UK AI Security Institute, home page. 5. GOV.UK. AI Security Institute, "Frontier AI Trends report factsheet". 6. UK AI Security Institute. "Pre-deployment evaluation of Anthropic's upgraded Claude 3.5 Sonnet", 19 November 2024 (joint with US AISI). · Last updated 21 November 2026. Every figure is footnoted to its source.

ONE YOUNG INDIA
ORGANISER

FLAME UNIVERSITY
HOST CAMPUS AND KNOWLEDGE PARTNER

AFS Intercultural Programs India
PROGRAMME PARTNER

SECTION 1

Letter from the chairs

Dear delegates,

We will chair the Digital Economy and Artificial Intelligence Working Group at the Model G20 Summit 2027. Our group asks one hard question: should G20 members agree common safety tests for the most powerful AI models before they are released?

Read this guide twice. The first time, read for the story of how governments began testing AI models, and why they disagree about who should test, when, and by whose rules. The second time, read it as your member's negotiator and mark every line that helps or hurts your case.

In the room we will look for clauses that name who acts, what they do and by when. We will also look for delegates who bring a member that wants binding rules and one that wants room to innovate to the same text, and who can show a source for every claim they make.

Bring your white paper, your member's row from Section 8 and one draft clause. We look forward to meeting you at FLAME University in June.

Warm regards,

Your chairs, Digital Economy and Artificial Intelligence Working Group

SECTION 2

How the G20 and this simulation work

The real G20

The Group of Twenty has 21 members: 19 countries and two regional bodies, the European Union and the African Union[7]. The African Union became a permanent member at the New Delhi summit in 2023[7]. Together the members account for around 85% of the world's economic output, over 75% of world trade and about two-thirds of its people[7].

The G20 began in 1999, after the Asian financial crisis, as a meeting of finance ministers and central bank governors[7]. Leaders first met in Washington DC in 2008, and they have met every year since 2011[7]. Each year one member holds the presidency and sets the agenda, helped by the members that held it before and will hold it after. These three are called the troika[7]. The United States took over the presidency on 1 December 2025 and cut the number of working groups from 16 to four[7]. The United Kingdom holds the presidency in 2027 and will host the Leaders' Summit in Manchester on 27 and 28 November 2027[8,9].

Two tracks

The G20 process is led by the sherpas, each leader's personal representative. The work runs in two tracks, the Finance Track and the Sherpa Track. Within the two tracks, working groups bring together officials from the members, invited countries and international organisations[7].

Agreement by everyone

In September 2026 G20 finance ministers issued their text as a Chair's Statement[10]. A note to it says it was agreed by all G20 members present except China, which objected to four of its paragraphs[10]. In this simulation a text is agreed only when no member objects; the Rules of Procedure 2027 explain how.

This simulation

The Model G20 Summit 2027 is set in 2027, the year the United Kingdom holds the G20 presidency. The Model G20 secretariat chairs every working group, and the UK delegation is an ordinary member. All 21 members are seated in every group. Each group also has four invited seats for international organisations or guest countries; they take part in the debate but are not members. Model G20 is organised by One Young India and is independent of the G20 and of every G20 government.

Each working group is modelled on a group that has met in the real G20 in recent years, so delegates learn how the forum works. Your group negotiates one short text on its agenda. On Sunday afternoon the working groups' texts come together in the communiqué, which the Leaders' Session adopts by consensus at the closing ceremony. The Rules of Procedure 2027 explain how debate, amendments and the final text work.

This working group

The working group is modelled on the work of the G20's digital economy ministers. Meeting in Maceió in 2024, under Brazil's presidency, they reaffirmed the G20 AI Principles[11]. In 2025 the leaders' declaration agreed under South Africa's presidency said that, for AI to be safe and trustworthy, issues including regulation, safety and human oversight "must be addressed"[12]. In 2026 the US presidency held an Innovation Ministerial in Chapel Hill, North Carolina[13], whose statement calls on members to "preserve national sovereignty in the governance of emerging technologies"[14]. The UK Prime Minister has said he will put AI at the heart of the UK's presidency in 2027[15].

SECTION 3

The issue explained

The most powerful models

The countries at the AI Safety Summit at Bletchley Park in 2023 described "frontier" AI as highly capable general-purpose models that can match or exceed the most advanced models of today[3]. Telling which models are the most powerful is not simple. The EU's AI Act uses the amount of computing power used to train a model as one way to estimate what it can do[2]. Above 10^{25} operations, it presumes a model has high-impact capabilities[2], and the Act's recitals say a model should be considered to present systemic risks if it has such capabilities or a significant impact on the EU market because of its reach[2].

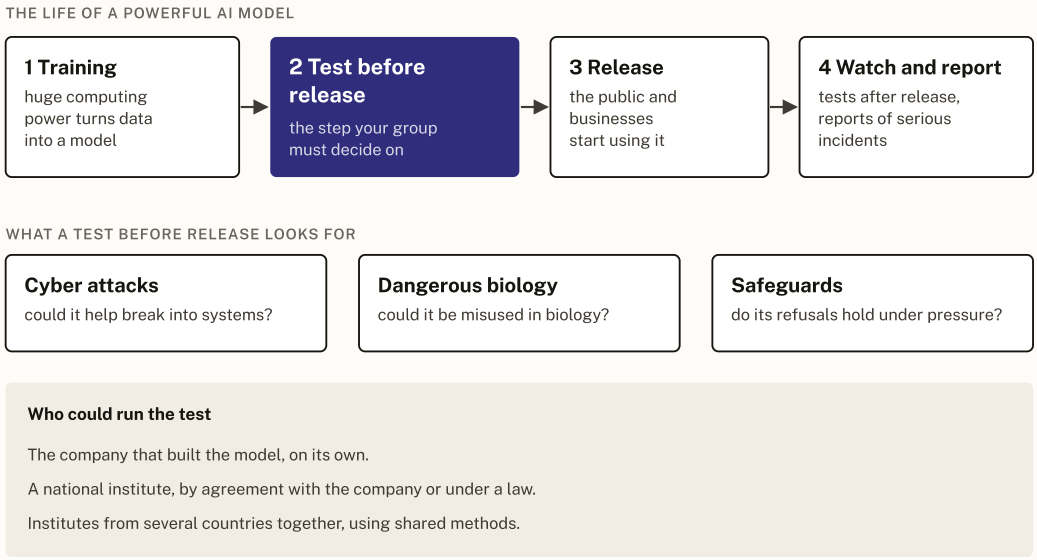
What a safety test does

The Bletchley Declaration says those building the most powerful AI systems have "a particularly strong responsibility" for their safety, "including through systems for safety testing, through evaluations, and by other appropriate measures"[3]. It names cybersecurity and biotechnology as areas of special concern[3]. Tests try to find out how far a model could help someone attack a computer system or misuse biology, and whether its safeguards hold when people try to get round them[6,16].

Governments have built teams to run these tests. The UK's AI Security Institute says it tests leading AI systems before and after they are released[5]. The EU requires the makers of the most powerful models to evaluate them using standardised methods, including adversarial testing, and to report serious incidents to its AI Office[2,17]. A test is not a verdict: in one joint UK and US evaluation, the institutes warned that the results of their safeguard tests "cannot on their own determine the model's risks"[6]. They called its findings preliminary, as they did in a second joint evaluation[6,18]. Figure 1 shows where testing fits in a model's life.

FIGURE 1 SECTION 3

Where safety testing fits in the life of an AI model, and who does it.



Why agreeing on tests is hard

Tests are hard to compare. When the US, UK and Singapore institutes first tested a model together, in 2024, they found that small differences in method "can have a large impact": on one test their results differed by more than 5 percentage points until they found the cause and aligned their results[19]. In their third joint exercise, in 2025, testing AI agents was so new that the main focus was on methods, not on results[20].

Members also take different routes. The EU, Korea and China have binding rules[2,21,22], while the US testing centre plans voluntary agreements[16]. Even so, the 2026 Innovation Ministerial statement, which the White House calls a consensus statement, warns: "Outdated or inflexible regulatory approaches can unintentionally restrict innovation"[13,14]. In February 2026 the United States registered its "strong objection" to the UN scientific panel on AI "as currently constituted"[23]. Writing about the new panel, UN News said that diverging attitudes towards regulation and growing geopolitical tensions could make consensus hard to forge[24].

SECTION 4

What has happened so far

30 Oct 2023	G7 leaders agree guiding principles and a voluntary code of conduct for AI developers[25]
1 to 2 Nov 2023	Countries at the AI Safety Summit issue the Bletchley Declaration[3]
21 May 2024	The AI Seoul Summit issues a statement on international cooperation on AI safety science[26]
13 Sep 2024	G20 digital economy ministers meet in Maceió, under Brazil's presidency[11]
21 Nov 2024	Ten participants launch the International Network of AI Safety Institutes in San Francisco[26]
10 to 11 Feb 2025	The AI Action Summit in Paris brings together participants from over 100 countries[27]
10 Jul 2025	The voluntary General-Purpose AI Code of Practice, prepared by independent experts, is published[28]
17 Jul 2025	The Network publishes the results of its third joint testing exercise[20]
2 Aug 2025	The EU AI Act's obligations for general-purpose AI models begin; enforcement powers follow from 2 August 2026[17]
22 Nov 2025	A leaders' declaration agreed under South Africa's presidency includes a section on AI[12]
12 Feb 2026	The UN General Assembly appoints the 40 members of its scientific panel on AI, by 117 votes to 2[23]
6 to 7 Jul 2026	The first UN Global Dialogue on AI Governance takes place[29]
2 Sep 2026	The G20 Innovation Ministerial in Chapel Hill ends with a statement[13]
22 Sep 2026	The UK Prime Minister tells the UN General Assembly he will put AI at the heart of the UK's 2027 G20 presidency[15]

What the G20 has already said

The leaders' declaration agreed under South Africa's presidency, Johannesburg, November 2025:

> *"To ensure safe, secure, and trustworthy AI development, deployment and use, the protection of human rights, transparency and explainability, fairness, accountability, regulation, safety, appropriate human oversight, ethics, biases, privacy, data protection and data governance must be addressed."*[12]

G20 Innovation Ministerial Statement, Chapel Hill, September 2026:

> *"We call on members to develop their own policies and preserve national sovereignty in the governance of emerging technologies, while allowing for engagement in international discussions."*[14]

SECTION 5

Facts that matter

FIGURE	VALUE	YEAR	REF.
Global AI market (UNCTAD projection)	US$189 billion to US$4.8 trillion	2023 to 2033	1
Share of world AI research and development by 100 companies	40%	2022	1
Jobs that AI could affect, UNCTAD	40% of jobs worldwide	2025 report	1
EU threshold for presuming high-impact capabilities	10^{25} operations used in training	from August 2025	2,17
Highest EU fine for makers of general-purpose AI models	3% of yearly worldwide turnover or €15 million, whichever is higher	from August 2026	2,17
Organisations signing the Frontier AI Safety Commitments	16, plus 4 added later	2024	30
Participants at the launch of the Network of AI Safety Institutes	10	2024	26
Public cyber challenges solved in the US institute's pre-release tests (40 in each test)	upgraded Claude 3.5 Sonnet 32.5%; o1 45%; best reference model 35%	2024	6,18
Safety tests on leaks of sensitive information and fraud passed by AI agents in English, third joint testing exercise	about 40% overall	2025	20
UN vote appointing the scientific panel on AI	117 in favour, 2 against	2026	23
Members of the UN scientific panel	40, 19 women and 21 men	2026	23,24
Delegations the White House lists as present at the 2026 G20 Innovation Ministerial, besides the US host	20, including Poland	2026	13

SECTION 6

Core tensions

Test before release, or keep room to innovate

For testing before release. The Bletchley Declaration gives the builders of the most powerful systems a particularly strong responsibility for safety, including through safety testing[3]. The EU AI Act's recitals, which explain its rules, say providers should perform the necessary model evaluations "in particular prior to its first placing on the market"[2]. The UK Prime Minister has promised to work towards "a single set of global principles and standards" so that AI development is safe[15].

For room to innovate. The 2026 Innovation Ministerial statement warns that inflexible rules can restrict innovation[14]. Even the Bletchley Declaration says countries should consider "a pro-innovation and proportionate governance and regulatory approach"[3].

Common tests, or national choices

For common tests. Small differences in method can have a large impact on results[19], so tests run separately are hard to compare. Safety institutes from several countries have already tested models together three times[20].

For national choices. The 2026 Innovation Ministerial statement calls on members to develop their own policies and keep national sovereignty over how emerging technologies are governed[14]. The Global AI Governance Action Plan also lists respecting national sovereignty among its principles[31]. It was published at the 2025 World AI Conference and posted by China's Foreign Ministry[31].

Voluntary, or required by law

For voluntary commitments. At the 2024 Seoul summit, companies agreed to assess the risks of their frontier models before deploying them, and to set thresholds at which severe risks, unless adequately mitigated, would be deemed intolerable[30]. The US government's testing centre says it will make voluntary agreements with AI developers[16].

For law. The EU AI Act requires model evaluation and adversarial testing, with fines of up to 3% of yearly worldwide turnover[2]. Korea's Framework Act on AI sets a duty to ensure safety for systems above a computing threshold[21], and China requires security assessments for some AI services[22].

SECTION 7

Case studies

The UK and US test models before release

On 19 November 2024 the UK and US AI safety institutes published a joint evaluation of an upgraded Claude 3.5 Sonnet model, made by Anthropic, carried out during a limited period before its release[6]. On 18 December 2024 they published a joint evaluation of OpenAI's o1 model, tested in three areas, including cybersecurity[18]. In each evaluation the US institute ran a suite of 40 public cybersecurity challenges: the upgraded Claude 3.5 Sonnet solved 32.5% and o1 solved 45%, against 35% for the best reference model in each[6,18]. In the Claude 3.5 Sonnet evaluation the institutes warned that their safeguard results "cannot on their own determine the model's risks"[6]. Both write-ups call their findings preliminary[6,18]. In 2026 the US testing centre also assessed models from the Chinese company Z.ai after they were released[16].

What it shows. Governments can test a model before release when its maker gives them access, and the results still need careful reading. **Question for your group.** Should every maker of the most powerful models be expected to give governments access before release, or only those that agree?

The EU's Code of Practice

The General-Purpose AI Code of Practice, prepared by independent experts, was published on 10 July 2025[28]. It has three chapters: transparency, copyright, and safety and security, and the safety chapter applies only to makers of the most advanced models[28]. The Commission describes that chapter as

setting out state-of-the-art practices for managing systemic risks[28]. Signatories set up a taskforce, chaired by the EU's AI Office, to apply the Code consistently[28]. The Commission's powers to enforce the AI Act's rules for these models began on 2 August 2026[17].

What it shows. A voluntary code can sit alongside binding law and explain how to meet it. **Question for your group.** Could a G20 text borrow this model: shared practices first, binding rules where members choose?

China's security assessments

Under China's 2023 interim measures (in an unofficial English translation by China Law Translate), providers of generative AI services with "public opinion properties" or the capacity for social mobilisation must carry out security assessments and file their algorithms[22]. The measures took effect on 15 August 2023[22]. They say the state places equal emphasis on development and security and carries out "tolerant and cautious graded management by category" of these services[22]. In 2025 the Global AI Governance Action Plan, published at a conference and posted by China's Foreign Ministry, said: "We need to explore categorized and tiered management approaches, build a risk testing and evaluation system for AI"[31].

What it shows. A large AI power can require security assessments at home while calling for respect for national sovereignty abroad. **Question for your group.** Can a G20 text on testing work for members whose systems differ this much?

SECTION 8

Country positions

One row for each of the 21 members. "Stake" and "On the record" come from official sources, except where a row names a broadcaster or a translation. "Our reading", in italics, is what we think that record suggests: test it in your research. Where the record is silent, the row says so. Signing the Bletchley Declaration or the Paris summit statement, or attending the 2026 Innovation Ministerial, is not a position on testing. The label under each name shows the member's place in the record: institute (it has a national AI safety or security institute or testing centre), binding rules (it has a law that requires testing or assessment), or not clear. Australia, Canada and France are labelled not clear, but they took part in the Network's third joint testing exercise[20]. Bring this section to the summit and tick each row as members speak.

MEMBER	STAKE	ON THE RECORD	OUR READING	MARK
Argentina NOT CLEAR	It attended the 2026 Innovation Ministerial[13], and is on neither the Bletchley nor the Paris summit list[3,27].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Australia NOT CLEAR	It was one of the ten participants at the launch of the International Network of AI Safety Institutes[26].	It took part in the Network's third joint testing exercise in July 2025[20], and is on the Bletchley and Paris summit lists[3,27].	***Pushes for:*** *Joint testing between institutes.* ***Allies:*** *Canada, Japan, the United Kingdom.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing

MEMBER	STAKE	ON THE RECORD	OUR READING	MARK
Brazil NOT CLEAR	Its Senate approved an AI bill in December 2024 and sent it to the Chamber of Deputies in March 2025[32].	The Maceió declaration, agreed under its 2024 presidency, reaffirmed the G20 AI Principles[11]. It is on the Bletchley and Paris summit lists[3,27].	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Canada NOT CLEAR	It was one of the ten participants at the launch of the Network of AI Safety Institutes[26].	The Network met in Vancouver in July 2025, alongside its third joint testing exercise[33]. It is on the Bletchley and Paris summit lists[3,27].	***Pushes for:*** *Joint testing between institutes.* ***Allies:*** *The United Kingdom, Japan, Australia.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing
China BINDING RULES	Its 2023 measures, in an unofficial English translation by China Law Translate, require some generative AI services to carry out security assessments and file their algorithms[22].	Its Foreign Ministry posted the Global AI Governance Action Plan, published at the 2025 World AI Conference, which calls on all parties to act on principles including respecting national sovereignty and ensuring safety[31]. The plan says: "We need to explore categorized and tiered management approaches, build a risk testing and evaluation system for AI"[31].	***Pushes for:*** *Testing under national rules.* *Not clear from the record: red line, allies. Research it.*	☐ Ally ☐ Opp. ☐ Swing
France NOT CLEAR	It hosted the AI Action Summit in Paris in February 2025[27], and took part in the Network's third joint testing exercise, with the European Commission and others[20].	The Paris summit statement says that harnessing the benefits of AI "depends on advancing Trust and Safety"[27].	***Pushes for:*** *Trust and safety alongside innovation.* ***Allies:*** *The European Union.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Germany INSTITUTE	Its government reports that a national AI safety institute, AISI Deutschland, has opened[34].	It is on the Bletchley and Paris summit lists[3,27]. No further public statement on testing found.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
India NOT CLEAR	It hosted the India AI Impact Summit in 2026; the national broadcaster reports that its declaration was endorsed by 88 countries and organisations[35].	The same report says the declaration holds that the benefits of AI must be equitably shared[35].	***Pushes for:*** *Sharing the benefits of AI.* *Not clear from the record: red line, allies. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Indonesia NOT CLEAR	It is on the Bletchley and Paris summit lists[3,27] and attended the 2026 Innovation Ministerial[13].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Italy NOT CLEAR	It is on the Bletchley and Paris summit lists[3,27] and attended the 2026 Innovation Ministerial[13].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Japan INSTITUTE	Ten ministries and five government-affiliated organisations set up the Japan AI Safety Institute in February 2024[36].	Its AI Promotion Act of 2025 calls for measures such as transparency in how AI is developed and used, because misuse could help crimes, leaks of personal information and copyright infringement[37]. It took part in the Network's third joint testing exercise[20].	***Pushes for:*** *Joint testing between institutes.* ***Allies:*** *The United Kingdom, Canada, Korea.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing

MEMBER	STAKE	ON THE RECORD	OUR READING	MARK
Republic of Korea BINDING RULES	Its Framework Act on AI requires operators of systems trained with computing power above a threshold set by decree to assess and reduce risks and to monitor safety incidents[21].	With the UK it announced the organisations that agreed the Frontier AI Safety Commitments at the 2024 Seoul summit[30].	***Pushes for:*** *Safety duties above a computing threshold.* ***Allies:*** *The European Union, Japan.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Mexico NOT CLEAR	It attended the 2026 Innovation Ministerial[13], and is on the Paris summit list but not the Bletchley Declaration list[3,27].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Russia NOT CLEAR	It attended the 2026 Innovation Ministerial[13], and is on neither the Bletchley nor the Paris summit list[3,27].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Saudi Arabia NOT CLEAR	It is on the Bletchley Declaration list[3] but not the Paris summit list[27], and attended the 2026 Innovation Ministerial[13].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
South Africa NOT CLEAR	It held the G20 presidency in 2025[12].	The declaration agreed under its presidency commended the presidency's leadership in advancing global discussions on AI[12] and welcomed the AI for Africa Initiative[12].	***Pushes for:*** *AI that serves Africa.* ***Allies:*** *The African Union.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing
Türkiye NOT CLEAR	It is on the Bletchley Declaration list[3] but not the Paris summit list[27], and attended the 2026 Innovation Ministerial[13].	No public statement found. Research it.	*Not clear from the record. Research it.*	☐ Ally ☐ Opp. ☐ Swing
United Kingdom INSTITUTE	It holds the G20 presidency in 2027[8]. Its AI Security Institute calls itself the first state-backed organisation dedicated to ensuring advanced AI is safe, secure and beneficial[4].	In September 2026 its Prime Minister promised to work towards "a single set of global principles and standards" and to put AI at the heart of its G20 presidency[15], and said the UK would bring forward new laws "if that is what is needed"[15]. With the US, its institute tested two models before their release in 2024[6,18].	***Pushes for:*** *Shared global standards for testing.* ***Allies:*** *Canada, Japan, France.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing
United States INSTITUTE	It held the G20 presidency in 2026[7] and hosted the Innovation Ministerial[13]. Its testing centre says it will make voluntary agreements with developers and lead evaluations of risks to national security[16].	In February 2026, citing sovereignty, it registered its "strong objection" to the UN scientific panel on AI "as currently constituted", voted against appointing its members and urged the UN not to "attempt to regulate or stifle the development of" cutting-edge technologies[23]. Its testing centre aims to guard against "burdensome and unnecessary regulation of American technologies by foreign governments" and to ensure US dominance of international standards[16].	***Pushes for:*** *Voluntary testing and national choice.* ***Red line:*** *A UN role in regulating AI.* *Not clear from the record: allies. Research it.*	☐ Ally ☐ Opp. ☐ Swing
European Union BINDING RULES	Its AI Act requires makers of general-purpose models with systemic risk to evaluate them, including with adversarial testing, and to report serious incidents to the AI Office[2].	Its enforcement powers over these rules began on 2 August 2026[17]. In 2023 the European Commission welcomed the G7's voluntary code of conduct for AI developers[25]. The EU is on the Bletchley and Paris summit lists[3,27].	***Pushes for:*** *Binding evaluation of the most powerful models.* ***Allies:*** *France, Germany, Korea.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing

MEMBER	STAKE	ON THE RECORD	OUR READING	MARK
AU **African Union** NOT CLEAR	The declaration agreed under South Africa's presidency in 2025 welcomed the AI for Africa Initiative, a voluntary platform for cooperation between the G20 and the African Union[12].	The Paris summit statement cites the African Union's Continental AI Strategy among existing initiatives on AI[27].	***Pushes for:*** *Help for Africa to build and use AI.* ***Allies:*** *South Africa.* *Not clear from the record: red line. Research it.*	☐ Ally ☐ Opp. ☐ Swing

Invited seats (not members)

OECD. Its AI Principles were adopted in May 2019 and updated in May 2024[38].

International Telecommunication Union. Its Secretary-General is Vice-Chair of the AI for Good Global Commission, which first met at the AI for Good Global Summit 2026, held back-to-back with the first UN Global Dialogue on AI Governance[29].

UNESCO. Its Recommendation on the Ethics of AI was adopted by all 193 member states in November 2021[39].

Singapore (guest country). One of the ten participants at the launch of the Network[26], and one of three institutes that led its first pilot test[19].

The camps

Members that test together

The United Kingdom, the United States, Japan, Canada, France and Australia; Korea and the European Commission also took part in joint tests[6,18,20].

Binding rules

The European Union, Korea and China[2,21,22].

National choice first

The Global AI Governance Action Plan posted by China's Foreign Ministry[31]. The 2026 Innovation Ministerial statement, which the White House calls a consensus statement, also calls on members to preserve national sovereignty[13,14].

The UN route

The declaration agreed under South Africa's presidency recognised the UN's role in AI cooperation[12]; the United States objected to the UN panel[23].

SECTION 9

Options on the table

Each option is something the G20 can actually do: commit its own members, call on others, or ask international bodies to act and report. Who may back or resist each option is our reading of the record in Section 8, for you to test in your research.

A. Voluntary commitments to test before release

What it is. Members ask the makers of the most powerful models to let national institutes test them before release, by agreement. **Who runs it.** National institutes and the companies. **Precedent.** The 2024 Frontier AI Safety Commitments, in which companies agreed to assess risks before deploying their models and to set thresholds at which severe risks, unless adequately mitigated, would be deemed intolerable[30], and the US testing centre's stated plan to establish voluntary agreements with developers[16]. **For.** It can start at once, without new laws. **Against.** A company that does not agree is not tested. **May back it.** The United States and the United Kingdom.

B. Joint testing between national institutes

What it is. Members' institutes agree shared methods and test models together, so that one test can count in many countries. **Who runs it.** The International Network of AI Safety Institutes[26]. **Precedent.** The Network ran three joint testing exercises between November 2024 and July 2025[20]. **For.** Shared methods make results comparable: in the first joint test, results that differed by more than 5 percentage points were aligned once the institutes found the cause together[19]. **Against.** Fewer than half of G20 members took part in the Network's launch[26]. **May back it.** Japan, Canada, France and Australia, all Network participants[26].

C. Binding law

What it is. Members require by law that the most powerful models are evaluated before release. **Who runs it.** Each member's regulator. **Precedent.** The EU AI Act, with fines of up to 3% of yearly worldwide turnover or €15 million, whichever is higher[2]; its recitals ask for evaluation "in particular prior to its first placing on the market"[2]. Korea's Framework Act also sets safety duties above a computing threshold[21]. **For.** It reaches every covered model sold in a member's market. **Against.** The 2026 Innovation Ministerial statement warns that inflexible rules can restrict innovation[14]. **May back it.** The European Union and Korea. **May resist it.** The United States.

D. Shared standards for testing

What it is. Members work through standards bodies on common methods for testing AI models. **Who runs it.** National and international standards bodies. **Precedent.** The 2026 Innovation Ministerial statement says technical standards and regulations should give investment certainty and avoid unnecessary obstacles to international trade[14]. **For.** Standards can be agreed without a new treaty. **Against.** Standards take time to agree, and models change fast. **May back it.** The United States and the United Kingdom[15,16].

E. Report serious incidents

What it is. Makers of the most powerful models report serious incidents to a national authority, and members share what they learn. **Who runs it.** National authorities, with the OECD tracking incidents[38]. **Precedent.** The EU AI Act requires makers of models with systemic risk to report serious incidents to the AI Office[2,17]. **For.** It adds evidence from real use after release; one pre-release evaluation warned that its safeguard results "cannot on their own determine the model's risks"[6]. **Against.** It acts after the harm. **May back it.** The European Union and Korea.

SECTION 10

What the communiqué must answer

1. Which models are covered: by computing power, by capability, or both?
2. What must a test check, and who runs it?
3. Before release, after, or both?
4. Are results shared between members, and with the public?
5. Voluntary, or required by law?
6. Who reports progress, to whom, and by when?

DRAFT PARAGRAPH, AS AMENDED — WEAK, THEN STRONG

> ~~We recognise the importance of safe, secure and trustworthy AI and encourage continued international discussion.~~ We commit that our national AI safety or security institutes will agree, by the end of 2028, a shared set of methods for testing the most powerful AI models before release, and will recognise each other's results where the same methods are used. We call on developers of these models to give our institutes access before release, and ask the OECD to report progress to our Digital Ministers by October 2028.

WHO ACTS	WHAT	FOR WHOM	HOW LONG	WHO REPORTS
G20 members' institutes, and AI developers	*shared test methods; mutual recognition; access before release*	*the public in every member*	*by the end of 2028*	*the OECD, by October 2028*

The weak paragraph names no one and promises nothing. The strong one commits only G20 members and calls on everyone else. It is an example, not the answer: your group may decide on a different course, or on none.

SECTION 11

Preparation

Your white paper

Write at most 1,000 words, with numbered sources, and send it to your chairs by Monday 31 May 2027 (Rules of Procedure 9.2). Use five parts:

1. Your member's stake, in three numbers with sources.
2. Its position on each question in Section 10.
3. Its red lines: what it will not accept.
4. One draft paragraph in the style of a communiqué, beginning "We commit" or "We call on".
5. Your sources, numbered, official ones first.

Five questions to research for your member

1. Does your member have a national AI safety or security institute, or a law on AI?
2. Did it sign the Bletchley Declaration, and is it in the Network of AI Safety Institutes?
3. Do companies based in your member build some of the most powerful models?
4. What did it say at the 2026 Innovation Ministerial or at the UN about AI?
5. Which members are most likely to agree with it, and which one in the other camp could it win over?

Starter sources

The Bletchley Declaration, reference 3

The G20 Innovation Ministerial Statement, 2026, reference 14

The 2025 declaration agreed under South Africa's presidency, reference 12

The UK AI Security Institute, reference 4

The US testing centre, reference 16

The EU's Code of Practice, reference 28

The third joint testing exercise, reference 20

UNCTAD on the AI market, reference 1

POCKET CARD

Your first hour as a delegate

- [ ] **Roll call.** Hold up your placard and answer "present" when your member is called.
- [ ] **Opening speech.** Give your member's stake in one number, its position, and one thing it wants in the text.
- [ ] **Listen.** As members speak, mark their rows in Section 8: ally, opponent or swing.
- [ ] **First caucus.** Find two allies, one in your camp and one in the other.
- [ ] **By the break.** Bring one clause that names who acts, what they do, and by when.

SECTION 12

Glossary

Adversarial testing
Testing a model by trying on purpose to make it misbehave, for example by getting round its safeguards[2,6].

AI safety institute
A government body that tests leading AI models and studies their risks[4,5].

AI agent
An AI system that can take actions on its own, such as browsing or running code.

Compute
The computing power used to train or run an AI model.

Frontier AI
Highly capable general-purpose AI models that can match or exceed today's most advanced models[3].

General-purpose AI model
A model that can do many different tasks, rather than one.

Incident
An event in which an AI system causes or nearly causes serious harm.

Mutual recognition
Members accepting each other's test results.

Pre-deployment testing
Testing an AI model before it is released to the public.

Safeguard
A measure that stops a model from being misused, such as refusing harmful requests.

Systemic risk
In EU law, the risk from a general-purpose AI model with high-impact capabilities or a large effect on the market[2].

Threshold
A level of capability or risk at which stronger rules apply.

References

All sources accessed 20 and 21 November 2026. Numbers in the text refer to this list.

1. UNCTAD. "AI market projected to hit $4.8 trillion by 2033, emerging as dominant frontier technology" (Technology and Innovation Report 2025). https://unctad.org/news/ai-market-projected-hit-48-trillion-2033-emerging-dominant-frontier-technology
2. European Union. Regulation (EU) 2024/1689 (Artificial Intelligence Act), Official Journal L, 2024, EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689
3. GOV.UK. "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023". https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
4. UK AI Security Institute, home page. https://www.aisi.gov.uk/
5. GOV.UK. AI Security Institute, "Frontier AI Trends report factsheet". https://www.gov.uk/government/publications/ai-security-institute-frontier-ai-trends-report-factsheet/ai-security-institute-frontier-ai-trends-report-factsheet
6. UK AI Security Institute. "Pre-deployment evaluation of Anthropic's upgraded Claude 3.5 Sonnet", 19 November 2024 (joint with US AISI). https://www.aisi.gov.uk/blog/pre-deployment-evaluation-of-anthropics-upgraded-claude-3-5-sonnet
7. Ministry of External Affairs, India. "Group of Twenty - G20" (brief), July 2026. https://www.mea.gov.in/Portal/ForeignRelation/G20_2026.pdf
8. Prime Minister's Office, GOV.UK. "Growth and opportunity set to be at the heart of UK-hosted G20", 22 November 2025. https://www.gov.uk/government/news/growth-and-opportunity-set-to-be-at-the-heart-of-uk-hosted-g20
9. Cabinet Office, GOV.UK. "Manchester to host world leaders at the 2027 G20 Summit", 23 September 2026. https://www.gov.uk/government/news/manchester-to-host-world-leaders-at-the-2027-g20-summit
10. US Department of the Treasury. "G20 Chair's Statement", Second Meeting of G20 Finance Ministers and Central Bank Governors, Asheville, 31 August and 1 September 2026 (press release sb0620). https://home.treasury.gov/news/press-releases/sb0620
11. G20 Ministers responsible for the Digital Economy. "G20 Ministerial Declaration", Maceió, 13 September 2024 (Brazilian presidency), copy on GOV.UK. https://www.gov.uk/government/publications/g20-ministerial-declaration-maceio-13-september-2024/g20-ministerial-declaration-13-september-2024
12. G20. "G20 South Africa Summit: Leaders' Declaration", Johannesburg, 22 November 2025 (South African presidency), copy hosted by the G20 Research Group, University of Toronto. https://g20.utoronto.ca/2025/251122-declaration.html
13. The White House. "G20 Innovation Ministerial Concludes with Consensus Statement", 2 September 2026. https://www.whitehouse.gov/releases/2026/09/g20-innovation-ministerial-concludes-with-consensus-statement/
14. G20 Ministers. "G20 Innovation Ministerial Statement", Chapel Hill, 2 September 2026 (US presidency), PDF published by the UK Government on GOV.UK. https://assets.publishing.service.gov.uk/media/6a996786f5b35599aec19180/g20-innovation-ministerial-statement.pdf
15. Prime Minister's Office, GOV.UK. "PM speech at UNGA: 22 September 2026" (transcript as delivered). https://www.gov.uk/government/speeches/pm-speech-at-unga-22-september-2026
16. NIST. Center for Advancing Innovation and Standards for Super Intelligence (CAISSI), also called the Center for AI Standards and Innovation (CAISI). https://www.nist.gov/caissi
17. European Commission. "Guidelines for providers of general-purpose AI models", Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/guidelines-gpai-providers
18. UK AI Security Institute. "Pre-Deployment evaluation of OpenAI's o1 model", 18 December 2024 (joint with US AISI). https://www.aisi.gov.uk/blog/pre-deployment-evaluation-of-openais-o1-model
19. NIST. "Improving International Testing of Foundation Models: A Pilot Testing Exercise from the International Network of AI Safety Institutes". https://www.nist.gov/document/aisi-pilot-testing-exercise-blog
20. UK AI Security Institute. "International joint testing exercise: Agentic testing", 17 July 2025. https://www.aisi.gov.uk/blog/international-joint-testing-exercise-agentic-testing
21. Korea Legislation Research Institute. "Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation for Trust" (English translation, Korea). https://elaw.klri.re.kr/eng_service/lawViewContent.do?hseq=73499
22. "Interim Measures for the Management of Generative Artificial Intelligence Services" (China, 2023), English translation by China Law Translate (unofficial), on Georgetown University's Emerging Technology Observatory, AGORA. https://agora.eto.tech/instrument/1137
23. United Nations Regional Information Centre. "General Assembly Appoints Artificial Intelligence Panel", UN press release. https://unric.org/en/general-assembly-appoints-artificial-intelligence-panel/
24. UN News. "AI 'moving at the speed of light' warns Guterres, unveiling recommendations for UN expert panel", 4 February 2026. https://news.un.org/en/story/2026/02/1166891
25. European Commission. "Hiroshima Process International Code of Conduct for Advanced AI Systems". https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems
26. Innovation, Science and Economic Development Canada. "The International Network of AI Safety Institutes: Mission statement", 21 November 2024. https://www.ised-isde.canada.ca/site/ised/en/international-network-ai-safety-institutes-mission-statement
27. Élysée (French Presidency). "Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet", AI Action Summit, Paris, 10 and 11 February 2025, posted 11 February 2025. https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet
28. European Commission. "The General-Purpose AI Code of Practice", Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai
29. ITU. "AI for Good and WSIS Forum 2026 close with global push for AI for everyone", press release, July 2026. https://www.itu.int/en/mediacentre/Pages/PR-2026-07-10-Geneva-Digital-Week-closing.aspx
30. GOV.UK. "Frontier AI Safety Commitments, AI Seoul Summit 2024". https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024
31. Ministry of Foreign Affairs of the People's Republic of China. "Global AI Governance Action Plan" (July 2025). https://www.mfa.gov.cn/mfa_eng/xw/zyxw/202507/t20250729_11679232.html
32. Federal Senate of Brazil. "PL 2338/2023" (bill on the use of artificial intelligence), the bill's progress page. https://www25.senado.leg.br/web/atividade/materias/-/materia/157233
33. European Commission. "AI Office contributes to the third joint testing exercise of the International Network of AI Safety Institutes", Shaping Europe's digital future, July 2025. https://digital-strategy.ec.europa.eu/en/news/ai-office-contributes-third-joint-testing-exercise-international-network-ai-safety-institutes
34. Die Bundesregierung (German federal government). "AISI Deutschland: KI-Sicherheitsinstitut eröffnet" (German). https://www.bundesregierung.de/breg-de/aktuelles/eroeffnung-ki-sicherheitsinstitut-2449410
35. Akashvani (All India Radio, Prasar Bharati), News On AIR. "India-AI Impact Summit concludes with adoption of New Delhi Declaration", 21 February 2026. https://www.newsonair.gov.in/india-ai-impact-summit-2026-adopts-new-delhi-declaration
36. Japan AI Safety Institute (J-AISI). "Factsheet of AI Safety in Japan 2025", 31 March 2026. https://aisi.go.jp/assets/pdf/j-aisi_factsheet_2025_en.pdf
37. Japan, Japanese Law Translation database. "Act on Promotion of Research and Development, and Utilization of Artificial Intelligence-Related Technology" (Act No. 53 of 4 June 2025), English translation. https://www.japaneselawtranslation.go.jp/en/laws/view/5066/en
38. OECD.AI. "AI Principles Overview". https://oecd.ai/en/ai-principles
39. UN News. "193 countries adopt first-ever global agreement on the Ethics of Artificial Intelligence", 25 November 2021. https://news.un.org/en/story/2021/11/1106612