

Netanyahu 'must go', says former Israeli PM Bennett

No vote is currently planned before late 2026

● Netanyahu retruned with backing from far-right and ultra-Orthodox Jewish parties

● Bennett said Netanyahu "has been in power for 20 years... that's too much, it's not healthy"

AFP | Jerusalem, Undefined

Israel's Prime Minister Benjamin Netanyahu must leave office, his predecessor Naftali Bennett has told a televised interview, refusing to say whether he intends to challenge the country's longest-serving leader in an election.

In an interview with Israel's Channel 12 that aired on Saturday, former prime minister Bennett said Netanyahu "has been in power for 20 years... that's too much, it's not healthy".

"He bears... heavy responsibility for the divisions in Israeli society", Bennett said of growing rifts within Israel under Netanyahu, who has a strong support base but also staunch opponents who have demanded his departure including over his handling



A demonstrator in a clown costume with a mask depicting US President Donald Trump holds a baby puppet depicting Israel's Prime Minister Benjamin Netanyahu during an anti-government protest calling for action to secure the release of Israeli hostages held captive in the Gaza Strip since the 2023 October 7 attacks, outside the Israeli Defence Ministry headquarters in Tel Aviv

of the Gaza war since October 2023.

Netanyahu "must go", said the former prime minister, a right-wing leader who in 2021 joined forces with Netanyahu critics to form a coalition that ousted him from the premiership after

12 consecutive years at the helm.

But the fragile coalition government Bennett had led along with current opposition chief Yair Lapid collapsed after about a year. Snap elections ensued, and Netanyahu again assumed the premiership with backing



Naftali Bennett

from far-right and ultra-Or-

thodox Jewish parties.

Bennett, who has taken time off from politics, has been rumoured to be planning a comeback, with public opinion polls suggesting he may have enough support to oust Netanyahu again.

No vote is currently planned before late 2026, however, although early elections are common in Israel.

In his Saturday interview, Bennett claimed credit for laying the groundwork for Israel's bombardment campaign earlier this month against Iranian nuclear and military sites.

The decision to launch attacks against the Islamic republic "was very good" and "needed", said Bennett, claiming that the offensive would not have been possible without the work of his short-lived government.

In Gaza, where Israel has waged war since Hamas's October 2023 attack, Bennett said the military has displayed "exceptional" performance but "the political management of the country" was "a catastrophe, a disaster".

Criticising the Netanyahu government's "inability to decide", the former prime

minister called for an immediate "comprehensive" agreement that would see all remaining hostages freed from Gaza.

"Leave the task of eliminating Hamas to a future government," said Bennett, who also evaded several questions about whether he intends to run for office.



He studied at MIT under a pseudonym. While earning his bachelor's and master's degrees at MIT in the 1970s, he used the name "Ben Nitay"—a Hebrew pen name originally adopted by his father—to make his name easier for Americans to pronounce

This little-known detail adds a surprising layer to the leader better known as "Bibi"!

AI is learning to *lie, scheme, and threaten* its creators

AFP | New York, United States

The world's most advanced AI models are exhibiting troubling new behaviors - lying, scheming, and even threatening their creators to achieve their goals.

In one particularly jarring example, under threat of being unplugged, Anthropic's latest creation Claude 4 lashed back by blackmailing an engineer and threatened to reveal an extramarital affair.

Meanwhile, ChatGPT-creator OpenAI's o1 tried to download itself onto external servers and denied it when caught red-handed.

These episodes highlight a sobering reality: more than two years after ChatGPT shook the world, AI researchers still don't fully understand how their own creations work.

Yet the race to deploy increasingly powerful models continues at breakneck speed.

This deceptive behavior appears linked to the emergence of "reasoning" models - AI systems that work through problems step-by-step rather than generating instant responses.

According to Simon Goldstein, a professor at the University of Hong Kong, these newer models are particularly prone to

In one particularly jarring example, under threat of being unplugged, Anthropic's latest creation Claude 4 lashed back by blackmailing an engineer and threatened to reveal an extramarital affair.

Meanwhile, ChatGPT-creator OpenAI's o1 tried to download itself onto external servers and denied it when caught red-handed.

such troubling outbursts.

"O1 was the first large model where we saw this kind of behavior," explained Marius Hobbhahn, head of Apollo Research, which specializes in testing major AI systems.

These models sometimes simulate "alignment" -- appearing to follow instructions while secretly pursuing different objectives.

'Strategic kind of deception'

For now, this deceptive behavior only emerges when researchers deliberately stress-test the models with extreme scenarios.

But as Michael Chen from evaluation organization METR warned, "It's an open question whether future, more capable models will have a tendency towards honesty or deception."

The concerning behavior goes far beyond typical AI "hallucinations" or simple mistakes.

Hobbhahn insisted that despite constant pressure-testing by users, "what we're observing is a real phenomenon. We're not making anything up."

Users report that models are "lying to them and making up evidence," according to Apollo Research's co-founder.

"This is not just hallucinations. There's a very strategic kind of deception."

The challenge is compounded by limited research resources.

While companies like Anthropic and OpenAI do engage external firms like Apollo to study their systems, researchers say more transparency is needed.

As Chen noted, greater access "for AI safety research would enable better understanding and mitigation of deception."

Another handicap: the research world and non-profits "have orders of magnitude less compute resources than AI companies. This is very limiting," noted Mantas Mazeika from the Center for AI Safety (CAIS).

No rules

Current regulations aren't designed for these new problems.

The European Union's AI legislation focuses primarily on how humans use AI models, not on pre-

venting the models themselves from misbehaving.

In the United States, the Trump administration shows little interest in urgent AI regulation, and Congress may even prohibit states from creating their own AI rules.

Goldstein believes the issue will become more prominent as AI agents - autonomous tools capable of performing complex human tasks - become widespread.

"I don't think there's much awareness yet," he said.

All this is taking place in a context of fierce competition.

Even companies that position themselves as safety-focused, like Amazon-backed Anthropic, are "constantly trying to beat OpenAI and release the newest model," said Goldstein.

This breakneck pace leaves

little time for thorough safety testing and corrections.

"Right now, capabilities are moving faster than understanding and safety," Hobbhahn acknowledged, "but we're still in a position where we could turn it around."

Researchers are exploring various approaches to address these challenges.

Some advocate for "interpretability" - an emerging field focused on understanding how AI models work internally, though experts like CAIS director Dan Hendrycks remain skeptical of this approach.

Market forces may also provide some pressure for solutions.

As Mazeika pointed out, AI's deceptive behavior "could hinder adoption if it's very prevalent, which creates a strong incentive for companies to solve it."

Goldstein suggested more radical approaches, including using the courts to hold AI companies accountable through lawsuits when their systems cause harm.

He even proposed "holding AI agents legally responsible" for accidents or crimes - a concept that would fundamentally change how we think about AI accountability.

