AIinLT 2026

Faculty of Education
CUHK 中大教育

AI in Language Education

Evidence · Paradigms · Implications for Hong Kong Schools

Alan Cheung

The Chinese University of Hong Kong
June 22, 2026

Blueprint for Digital Education Development in Primary and Secondary Schools

The DEBATE

THE OPTIMISTS

> *…to be as good a tutor as any human ever could.*
>
> **Bill Gates**

> *A personal tutor for every student; a teaching assistant for every teacher.*
>
> **Salman Khan, Khan Academy**

vs

THE SKEPTICS

> *ChatGPT is, basically, high-tech plagiarism.*
>
> **Noam Chomsky, linguist**

> *Language models are “stochastic parrots” — they mimic form without meaning.*
>
> **Emily M. Bender, computational linguist**

When the celebrities disagree, we don’t take sides. *We look at the evidence.*

CONTENTS

Part 1
Landscape as visible in the summit

Part 2
Evidence and paradigms

- **Chatbot-assisted language learning**
- **Robots' social behaviors for language learning**
- **Language development in human-AI interaction**

Part 3
Implications for Hong Kong schools

PART ONE

The landscape as visible in the summit

1

OPENING

One journey, a dozen names

Across English, Cantonese, and Putonghua, this summit kept reflecting the same shift.

- **From dependence to co-creation**
  從依賴到賦能共創
- **From handrail to letting go**
  從「扶手」到「放手」
- **From teacher-led to student-driven**
- **Writing with AI, not by AI**
- **From knowledge transmission to critical guidance**
  從知識傳遞到批判引導
- **From one-way generation to two-way dialectical engagement**
  從單向生成到雙向辯證
- **From consumption to creation**
- **From rote memory to deep immersion**
  從「硬背」到「深遊」
- **Reconfiguring agency & knowing**

The question has changed, not whether AI should be used, but how, for whom, and under what conditions.

From adoption to design

OLD QUESTIONS		NEW QUESTIONS
Can AI be used?		Under what conditions does it support learning?
Which tool is best?	DESIGN	Which pedagogical role should AI take?
Can it improve performance?		For which learners, skills, and contexts?
		How do we cultivate agency and responsibility?
		How do we stop misuse?

To answer these questions, we need two things: evidence and paradigms.

From tool to environment

Input
Texts, listening, examples

Interaction
Dialogue, role-play

Output
Writing, speaking, multimodal composing

Feedback
Grammar, vocabulary, pronunciation

Assessment
Scoring, diagnosis, progress

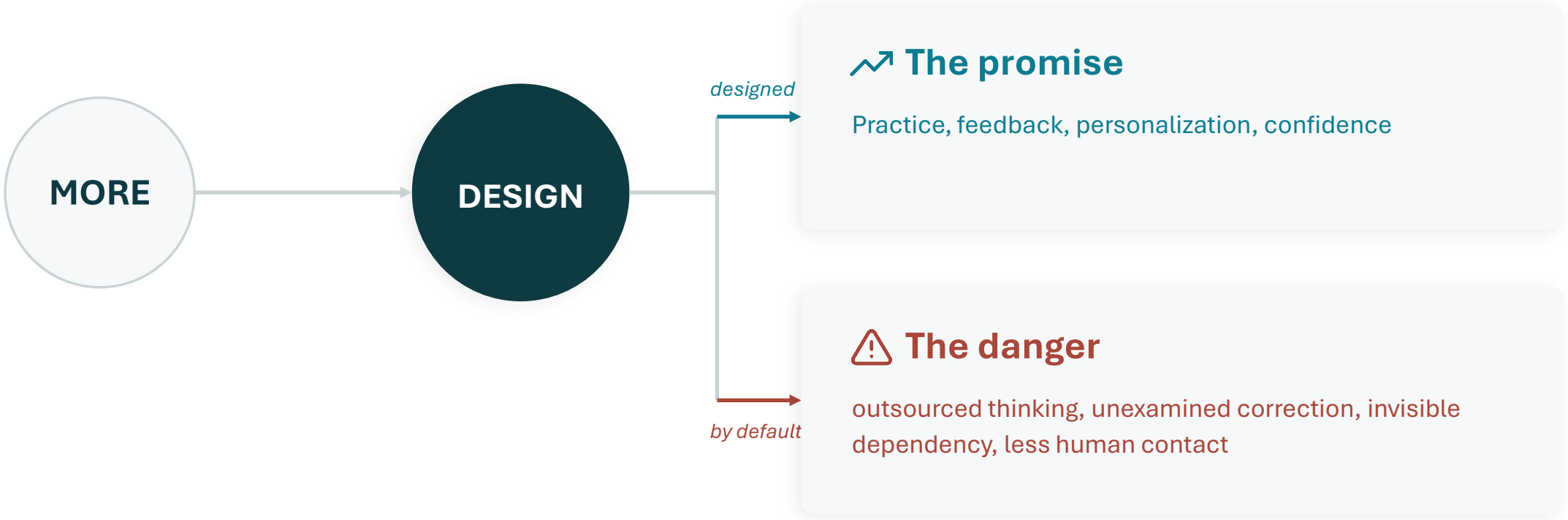
Reflection
Revision logs, metacognition

AI is entering the core processes of language learning.

A school without teachers

The promise and the danger

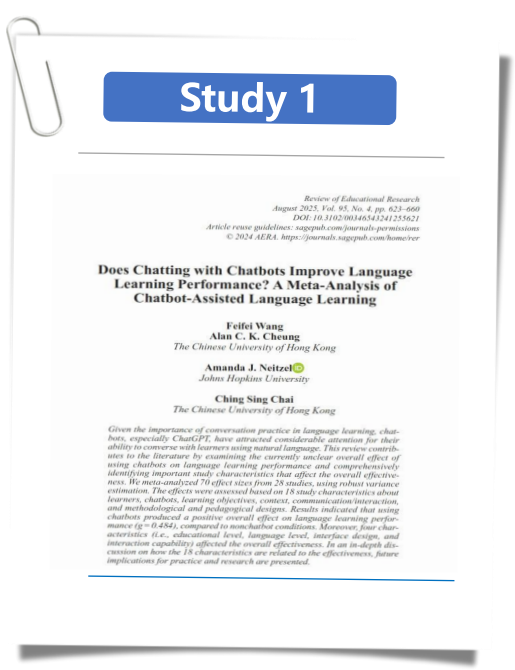
AI gives “more” practice, more feedback, more interaction.

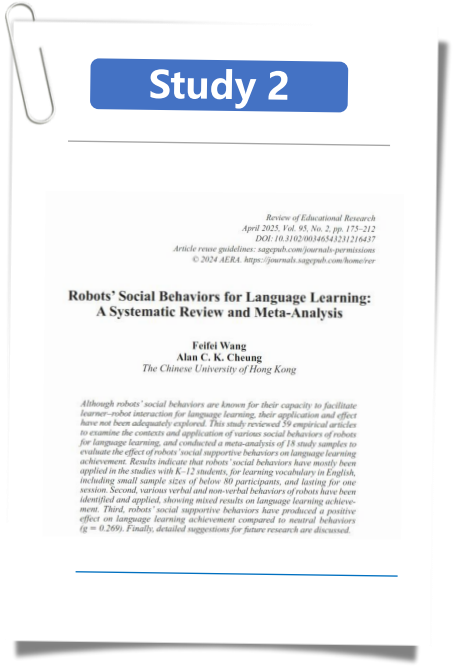


“More” is not automatically better, more should be designed.

PART TWO

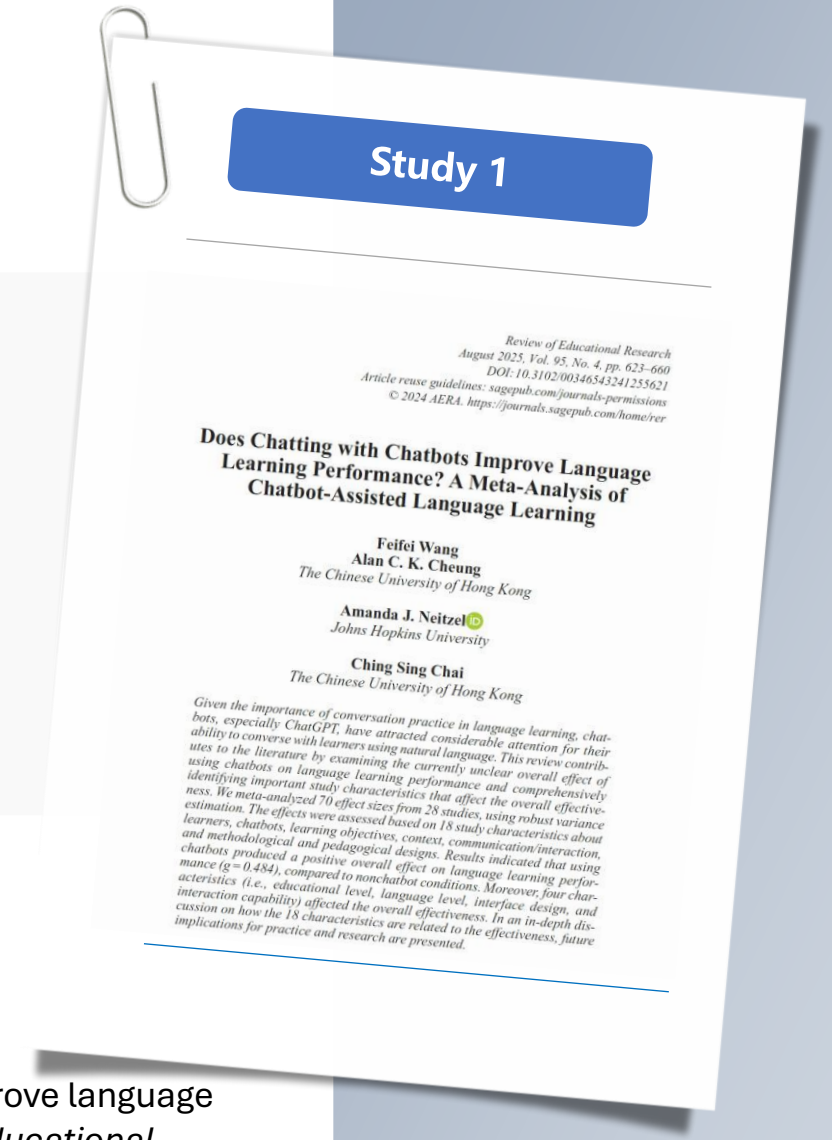
Evidence and paradigms: Three reviews





Review 1

Does chatting with chatbots improve language learning performance: Evidence from a meta-analysis



Wang, F., Cheung, A. C., Neitzel, A. J., & Chai, C. S. (2025). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research*, *95*(4), 623–660. https://doi.org/10.3102/00346543241255621

Research Questions

1. Does chatbot use improve language learning performance compared with non-chatbot conditions?

2. Not only “does it work?” but “under what conditions?”

Evidence base: A rigorous meta-analysis

5,478
records screened

→

239
full texts assessed

→

28
studies included

→

70
effect sizes extracted

2,641
learners in total

18
study characteristics

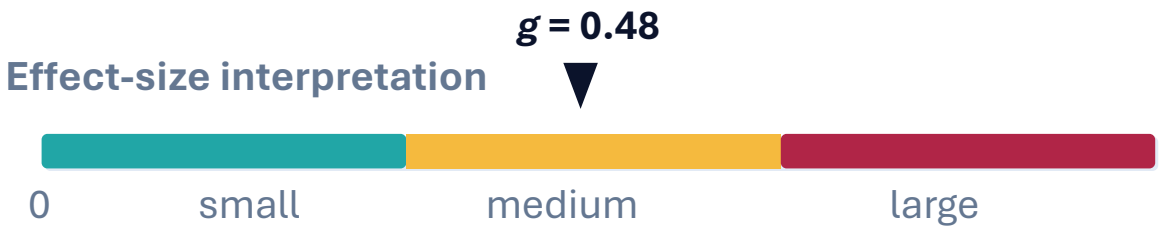
Key findings: A moderate positive effect

$g = 0.48$

95% CI [0.22, 0.75]

$p = .001$

Chatbot-assisted learning significantly improved language learning performance compared with non-chatbot conditions.



Conditional

Considerable heterogeneity means effectiveness may depend on learners, chatbot design, and implementation conditions.

54 positive
effect sizes

16 negative
effect sizes

Medium
overall effect

Key message: chatbots are promising, but their value is conditional.

For whom and with what designs?

Four learner–chatbot characteristics accounted for much of the variation in effectiveness.

Educational level

Positive effects for secondary students and adults. Very few elementary studies

Language level

Intermediate learners benefited more than beginners.

Interface design

Mobile-based interfaces showed a larger effect than web-based interfaces.

Interaction capability

Chatbot-driven scaffolding seemed more effective than user-driven open interaction.

Interpretation: The “right” chatbot is not universal. It should be matched to learner level, interface, and degree of scaffolding.

Design signals: Where effects looked strongest

Several positive adjusted mean effects highlight promising design choices for language classrooms.

Language target & interaction

- Vocabulary — $g = 0.881$
- Voice + text interaction — $g = 0.680$
- Individual interaction — $g = 0.528$
- L2 learning — $g = 0.556$

Classroom and pedagogical design

- In-class use — $g = 0.486$
- Teacher as facilitator — $g = 0.445$
- Chatbot's role of resource provider — $g = 0.475$
- Pedagogical principle of personalized learning — $g = 0.470$

Important caveat

These are design signals from adjusted mean effects. Not every signal was confirmed as an independent moderator.

Practical implication

Connect chatbots to specific language targets, guided interaction, and teacher mediation.

Reflection 1: Interaction needs support.

Chatbots work well as structured interaction partners.

Human first	AI in the middle	Human again
Set purpose	Dialogue practice	Evaluate suggestions
Initial output	Feedback / resources	Revise with judgment
Learner goal	Scaffolded revision	Transfer to real communication

Design questions for teachers

- **Which learner?**
- **Which target language skill?**
- **Which chatbot role?**
- **Which teacher mediation?**
- **What evidence of learning?**

The chatbot is not the endpoint of the task. It is in the middle of the task.

The Khanmigo lesson: Access is not interaction. AI tutors need learning design.

CHALKBEAT HOME COMMUNITIES TOPICS JOBS BOARD EVENTS ABOUT US

CHALKBEAT IDEAS AI IN EDUCATION TEACHING & CLASSROOM

Why Sal Khan's AI revolution hasn't happened yet, according to Sal Khan

By Matt Barnum | April 10, 2026, 5:11am GMT+8

REPUBLISH

Khan Academy CEO Sal Khan, center, and Microsoft Chief Technology Officer and Executive Vice President of Artificial Intelligence Kevin Scott speak at the Microsoft Briefing event at the Seattle Convention Center Summit Building in Seattle, Washington, on May 21, 2024. | (AFP via Getty Images)

Chalkbeat Ideas is a new section featuring reported columns on the big ideas and debates shaping American schools. Sign up for the Ideas newsletter to follow our work.

Three years ago, as Khan Academy founder Sal Khan rolled out an AI-powered tutoring chatbot, he predicted a revolution in learning.

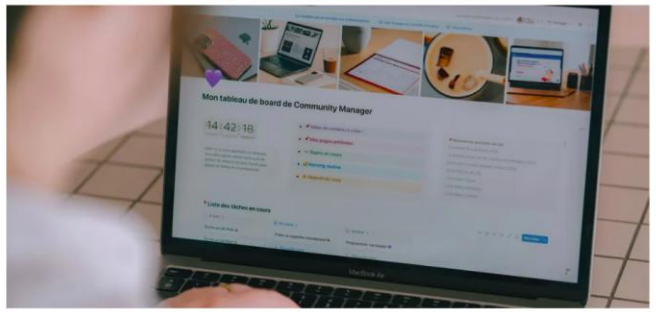
So far, the revolution hasn't happened, he acknowledges.

"For a lot of students, it was a non-event," Khan told me recently about his eponymous chatbot, Khanmigo. "They just didn't use it much."

Khan gives this analogy: Imagine he walked into a class, sat in the back of the room, and waited for students to seek out help. "Some will; most won't," he said. That's been the experience with AI tutoring, he said. It doesn't necessarily make students motivated to learn or fill in gaps in knowledge needed to ask questions.

Khan's comments are an acknowledgement that AI has not quickly allowed for the creation of an effective

DESIGN THINKING POW WOW 登入



設計思考 Design Thinking 2026-04-17 787

一個 AI 教育工具失敗之後，他做了一件更大的事

讀完這篇，你會知道：

1. **Khan Academy 創辦人 Sal Khan 公開承認，三年前推出的 AI 家教聊天機器人 Khanmigo 對多數學生而言「根本沒發生過什麼事」。**
2. **就在承認失敗的幾天後，他在 TED2026 發布了一個更大膽的計畫：聯合 TED、ETS，並與 Google、Microsoft、McKinsey 等企業合作，打造一所全新的大學—Khan TED Institute。**
3. **這個故事最值得思考的部分，不是這所學校會不會成功，而是「承認原型失敗、重新框架問題」這個動作本身，讓我們看到了什麼。**

思考的起點

把 AI 疊加在舊有的學習結構上面，和重新設計學習結構本身，是完全不同的兩件事。

2023 年，Sal Khan 在 TED 演講台上宣告：AI 即將帶來「教育史上最大規模的正面變革」，他承諾要讓地球上每一個學生都擁有一位「智慧型私人家教」。這場演講廣泛流傳，成為許多人對 AI 教育革命的想像基礎。

三年後，同樣是在 TED 的舞台旁，他說了另一句話。

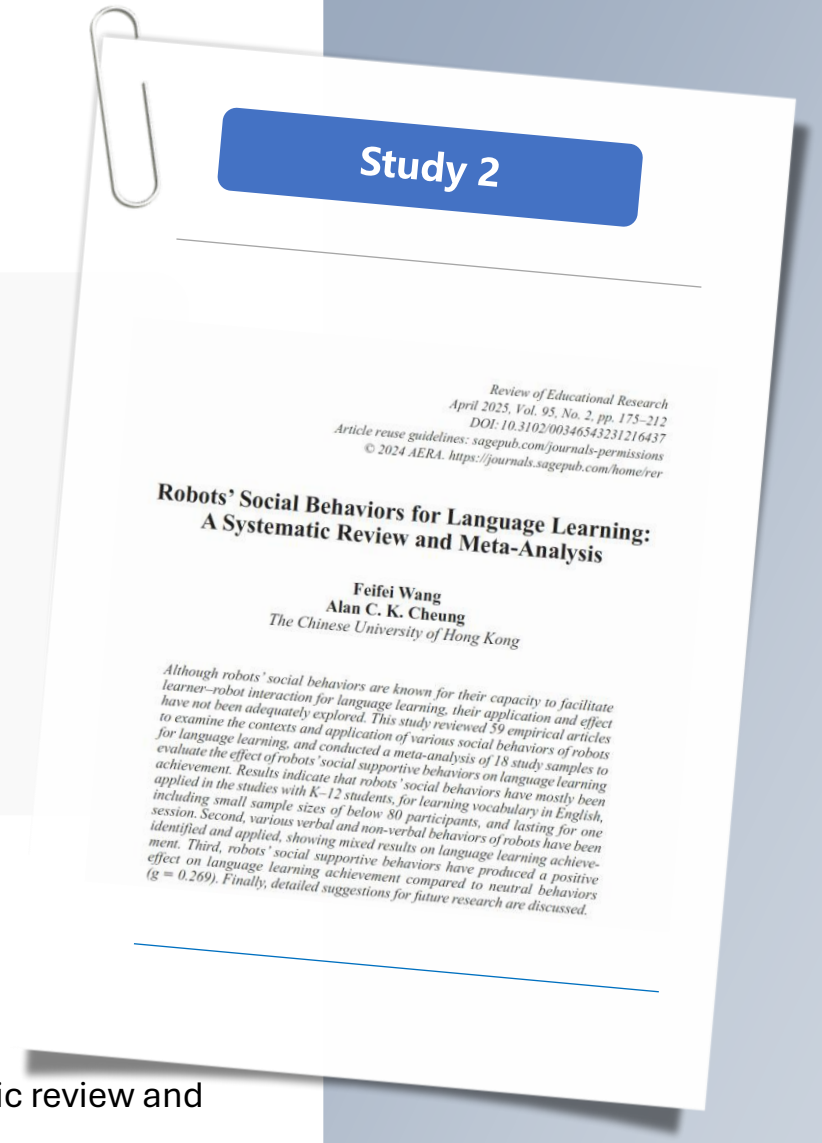
「對很多學生來說，那根本沒發生過什麼事。他們就是沒怎麼用它。」

他說的是 Khanmigo—那個曾被寄予厚望的 AI 家教。然後，他在同一週宣布了 Khan

- https://www.chalkbeat.org/2026/04/09/sal-khan-reflects-on-ai-in-schools-and-khanmigo/
- https://www.designthinking.com.tw/posts/942f9f86-a956-48cb-b4dd-1bb9b57545dd

Review 2

Robots' social behaviors for language learning: Evidence from a systematic review and meta-analysis



Wang, F., & Cheung, A. C. (2025). Robots' social behaviors for language learning: A systematic review and meta-analysis. *Review of Educational Research, 95*(2), 175–212.
https://doi.org/10.3102/00346543231216437

Robots' social behaviors for language learning

RQ1: Where and with whom are robots being used for language learning?

RQ2: Which verbal and non-verbal behaviors have been deployed?

RQ3: Do socially supportive behaviors produce better outcomes than neutral behaviors?

59
empirical articles

2,835
learners in the studies

Why this study matters

It shifts attention from "robot as device" to "robot as social actor with pedagogical roles."

What counts as robots' social behavior?

Social support is not simply "being cute." It is a designed combination of verbal and non-verbal signals.

Non-verbal behaviors

Physical	Physical appearance: size, form, body shape
Gesture	Gesture & posture: hand gestures, movement, actions
Face	Face & eye behaviors: facial expression, gaze, attention
Environment	Environment & space: distance, seating, classroom setup

Verbal behaviors

Vocal behavior	Vocal behaviors: voice identity, expressiveness
Verbal immediacy	Verbal immediacy: names, "we," self-disclosure
Verbal feedback	Verbal feedback: affective praise, corrective hints
Verbal adaptivity	Verbal adaptivity: level, needs, culture

Pedagogical interpretation

Robots become educationally meaningful when these social signals scaffold language use rather than distract from it.

Key findings: Small but positive effect

$g = 0.27$

95% CI [0.004, 0.533]

$p = .047$

Socially supportive robot behaviors improved language achievement compared with neutral behaviors.

18 samples
between-group comparisons

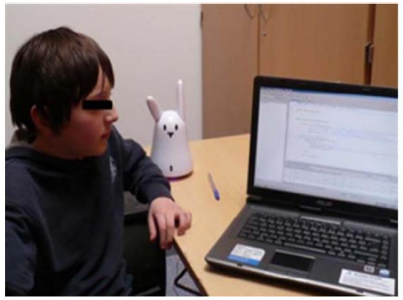
712 learners
meta-analytic sample

$I^2 = 65.578\%$
moderate heterogeneity

Key message: Social behaviors matter, but they are not a magic ingredient.

Role and social design are the leverage points

The effect was not uniform. The strongest signals came from which role the robot played and how social supportive behaviors were designed.

1. Face-to-face spatial arrangement in one-on-one interaction (Meiirbekov et al., 2016).	2. Side-by-side spatial arrangement in one-on-one interaction (Eimler et al., 2010).
	
3. Teacher-student gathering in group interaction (Hong et al., 2016).	4. Common-focus gathering in group interaction (Wu et al., 2015).
	

Role of robots

Peer

$g = -0.037$

5% CI [–0.417, 0.343]

vs.

Tutor

$g = 0.468$

95% CI [0.153, 0.783]

significant role difference

Pedagogical signal

Robots seemed more effective when positioned as tutors rather than peers.

Role and social design are the leverage points

The effect was not uniform. The strongest signals came from which role the robot played and how social supportive behaviors were designed.

Type of social supportive behavior

Verbal only	$g = 0.02$
Non-verbal	$g = 0.48$
Mixed	$g = 0.73$

significant type difference

Design signal

Combining verbal and non-verbal support appeared stronger than verbal or non-verbal support alone.

Reflection 2: Social supportive behaviors with pedagogical purposes

A robot that smiles is not yet a pedagogy. Social supportive behaviors become educational when it scaffolds language use.

Position the role

Use tutor-like support when the goal is explicit language learning.

→

Design the support

Combine encouragement, corrective feedback, gestures, and adaptation purposefully.

→

Control the novelty

Plan repeated use, teacher mediation, and transfer to human communication.

Social supportive behaviors should be tied to instructional purposes, not merely entertainment or novelty.

Review 3

Language learning development in human-AI interaction: A thematic review of the research landscape

Wang, F., Cheung, A. C., & Chai, C. S. (2024). Language learning development in human-AI interaction: A thematic review of the research landscape. System, 125, 103424. https://doi.org/10.1016/j.system.2024.103424

A thematic review of language learning development in human-AI interaction

6,540
Records retrieved from six databases

5,491
Records after duplicates removed

169
Full texts screened

49
Eligible empirical studies included

1 **Research question 1**

What are the contextual characteristics of research on language learning development in human-AI interaction?

2 **Research question 2**

What research paradigms of language learning development can be identified in human-AI interaction?

Paradigms were induced through eight constructs.

- Human–AI relationship
- Learning objective
- Task type
- Level of pre-structuring
- Mode of engagement behavior
- Knowledge-change process
- Cognitive outcome
- Research focus

Three paradigms emerged from the review

The paradigms are not simply three tools. They are three ways of distributing control among AI, teachers, and learners.

1

AI-directed, teacher-as-facilitator, learner-as-recipient

Human-AI relationship:
Stimulus–response

Learning objective:
To measure learning

Task type:
Closed-ended, high pre-structuring

Research focus:
To examine the effectiveness

32 studies

2

AI/teacher co-directed, learner-as-collaborator

Human-AI relationship:
Human–AI collaboration

Learning objective:
To foster learner-centered adaptive learning

Task type:
Semi-open, moderate pre-structuring

Research focus:
To understand and interpret the process

12 studies

3

AI/teacher/learner co-directed

Human-AI relationship:
Human–AI teaming

Learning objective:
Learner agency

Task type:
Open-ended, low pre-structuring

Research focus:
To foster learner agency

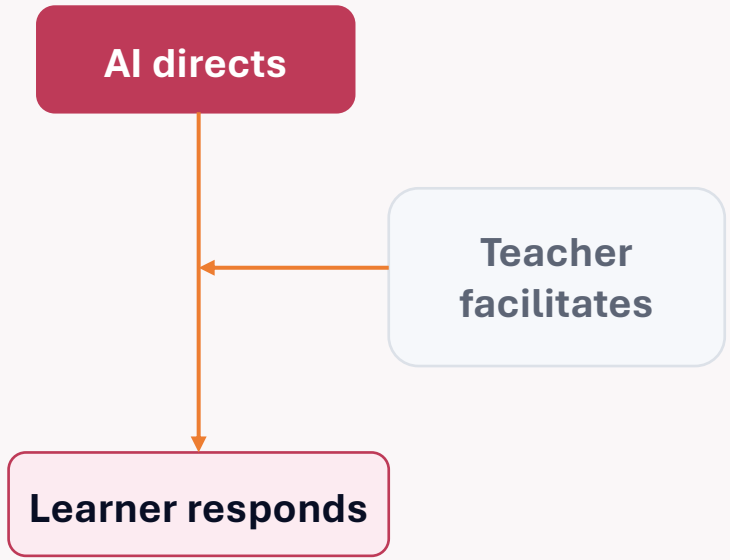
5 studies

A shift is visible: From AI as a controller of learning to AI as part of a human-centered learning relationship.

Paradigm 1: AI-directed, teacher-as-facilitator, learner-as-recipient

Useful when language goals are bounded, repeated, and measurable, but limited if it narrows learning to what AI can score.

Interaction pattern



Typical use

Automated scoring, vocabulary drills, pronunciation diagnosis, scripted robot interaction

Learning logic

High pre-structuring, closed tasks, fixed pathways, stimulus-response

Theoretical underpinnings

Positivism,
Universal Grammar Theory

Cognitive outcome

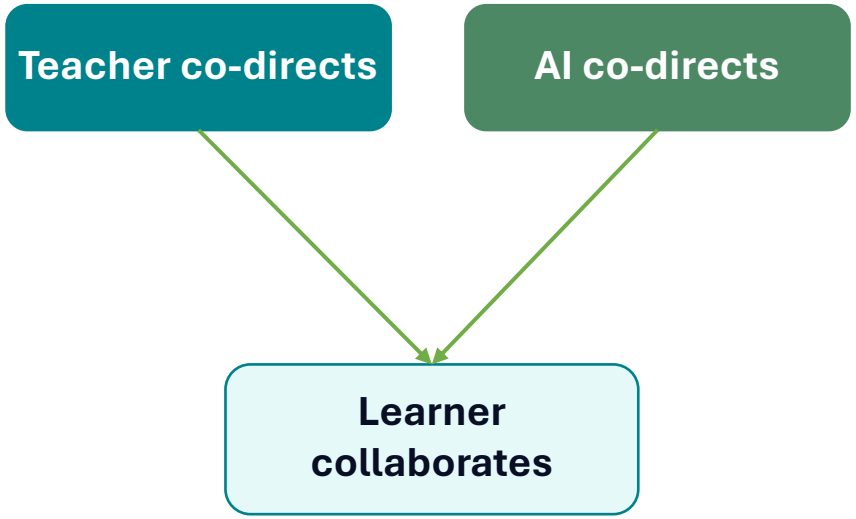
Recall or apply new knowledge to similar contexts

Takeaway

Use this paradigm for bounded practice and diagnosis, but avoid reducing language learning to measurable fragments.

Paradigm 2: AI/teacher co-directed, learner-as-collaborator

Most relevant for classroom integration: AI and teacher share direction, while learners collaborate and act on feedback.



Typical use

Semi-open tasks: chat, storytelling, guided revision, role-play

Learning logic

Moderate pre-structuring, semi-open-ended tasks, collaboration

Theoretical underpinnings

Phenomenology, constructivism

Cognitive outcome

Transfer new knowledge to novel contexts

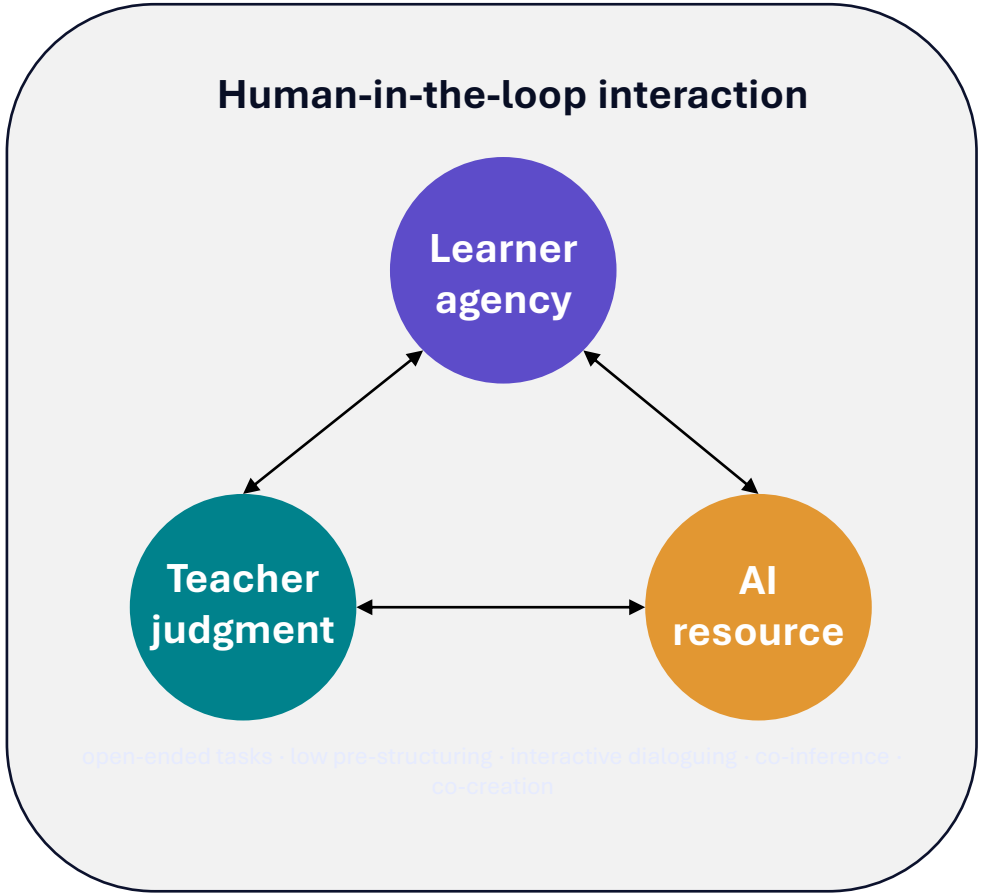
Teacher designs → AI scaffolds → Learner decides → Teacher consolidates

Takeaway

This is the practical classroom bridge between tool use and human-centered pedagogical design.

Paradigm 3: AI/teacher/learner co-directed

The agency-oriented paradigm places learners in the human loop: planning, monitoring, evaluating, and co-creating language.



Learner role

Planner, forethinker, self-regulator, self-examiner

Learning logic

Low pre-structuring, open-ended tasks, teaming

Theoretical underpinnings

Critical self-reflection, Dynamic Systems Theory

Cognitive outcome

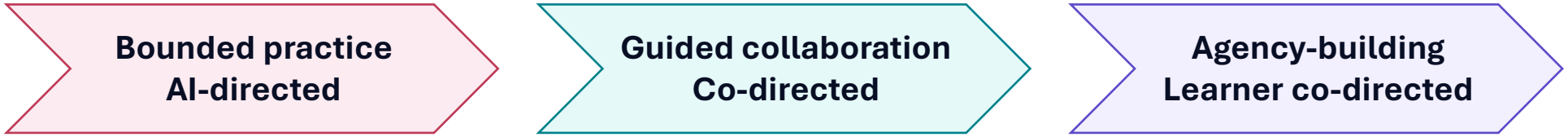
Co-inferring new knowledge

Takeaway

Not AI-assisted output, but AI-literate language agency

Reflection 3: Paradigms are design choices.

The review argues that the three paradigms differ in assumptions, but they can supplement each other in practice.



When useful

Clear targets, structured practice, diagnosis

When useful

Classroom tasks, writing revision, speaking rehearsal, reading, discussion

When useful

Open-ended inquiry, assessment redesign, AI literacy, identity and voice

Takeaway

- **In practice, the boundaries are blurry in practice, as both continuities and ruptures mark the boundaries among paradigms.**
- **Different paradigms reveal part of reality. The key is to ask which question is most significant to solve.**

Synthesis of the three reviews

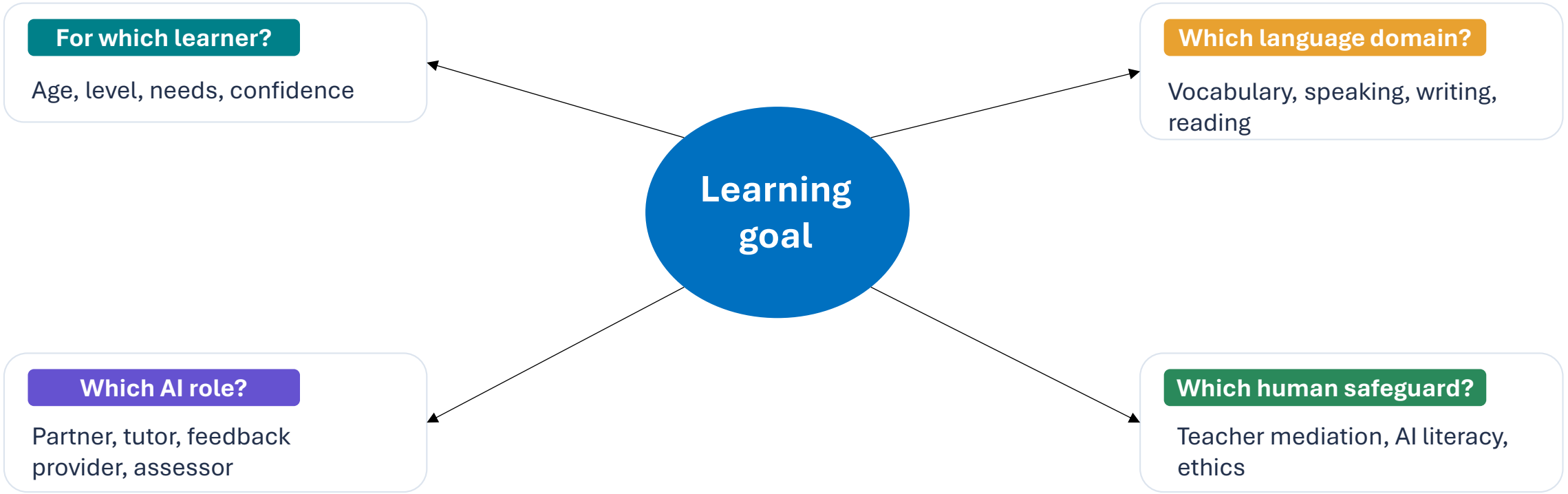
They shift the question from “Does AI work?” to “When, for whom, and in what role?”

Study 1 Chatbots	Interaction may improve language learning, especially when dialogue is structured and learners can act on feedback.
Study 2 Robots	Social supportive behaviors may help, but only when they serve a pedagogical role.
Study 3 Human-AI interaction paradigms	AI integration is a design choice: AI-directed, co-directed, or agency-oriented.

The key is not adoption. It is pedagogical design.

A design logic for AI in language education

Start with the learning goal, then ask four linked design questions.



Technology → Role → Task → Human mediation → Evidence

PART THREE

Implications for Hong Kong schools

3

Why Hong Kong is a high-stakes case?

AI integration should serve a multilingual and exam-sensitive system.

- **Biliteracy & trilingualism**

  Chinese + English literacy; Cantonese, Putonghua, and English communication

- **Learner diversity**

  Non-Chinese speaking students, SEN/dyslexia, various backgrounds, different home languages

- **Assessment pressure**

  HKDSE speaking and writing shape classroom priorities and evidence of learning

- **Policy momentum**

  AI for ALL subjects, school-based funding, subject-specific professional learning

From evidence to implications: Three design moves

1 Structure interaction

Chatbots help when dialogue has a purpose, scaffolding, and follow-up.

Study 1: Chatbots

2 Assign pedagogical roles

Social support matters when it helps learners practice, revise, and transfer.

Study 2: Robots

3 Build learner agency

Human-AI interaction varies from AI-directed to co-directed and agency-oriented.

Study 3: Paradigms

Implication for schools: Choose AI roles by learning goals, not by tool availability.

Start with Hong Kong language-learning priorities

AI use should begin with the learning problem.

Classroom challenge	Possible AI role	Preferred paradigm	Evidence echo
Limited oral practice for HKDSE / English	Speaking rehearsal partner + transcript	Co-directed	Structured interaction
Writing feedback workload and uptake	First-round feedback + revision coach	Co-directed	Teacher mediation
NCS / Putonghua / Cantonese pathways	Differentiated input + situational dialogue	Co-directed → agency	Learner differentiation
Reading motivation & vocabulary recycling	Reading companion + vocabulary recycler	AI-directed + co-directed	Vocabulary effects
Overreliance on AI-generated text	AI literacy	Agency-oriented	Human judgment

Question for practices: Which learner, which language domain, which AI role, and which safeguard?

Wrap-up 1: A classroom design pattern (human-AI-human)

AI should not replace the first act of thinking or the final act of communication.

Human first

Students: set purpose, draft / rehearse, predict difficulties, etc.

Teachers: frame the task

Set the goal & evaluation criteria. Decide what should be prepared through before AI.

AI in the middle

Students: role-play, get feedback / resources, compare models, etc.

Teachers: set the rules of engagement

Choose the tool & prompts. Define what AI may and may not do. Differentiate by level.

Human again

Students: evaluate suggestions, revise with judgment, etc.

Teachers: design the checkpoint

Oral, peer, or rubric. Assess the decisions and process, not just the product.

- **Teachers design the "before," manage the "middle," and assess the "after."**
- **Students remain the planner, evaluator, and communicator.**

Wrap-up 2: Assessment (from final product to evidence of decision-making)

AI-era assessment should reveal how learners think, revise, and transfer.

The risk

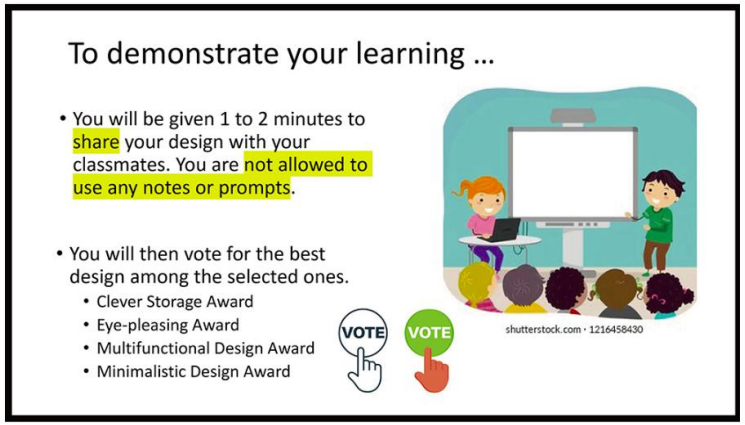
If we assess only the final text or score, AI makes learning less visible.

- outsourced thinking
- unexamined thinking
- generic voice
- reliance

Assess not only the final text, but also learners' decision-making behind the text.

An AI-visible assessment protocol

1	**Disclose**	What AI was used?
2	**Explain**	Which suggestions were accepted, rejected, or modified?
3	**Demonstrate**	Can the learner perform orally or in class?
4	**Transfer**	Can the strategy work in a new task?



Wrap-up 3: Teacher professional learning

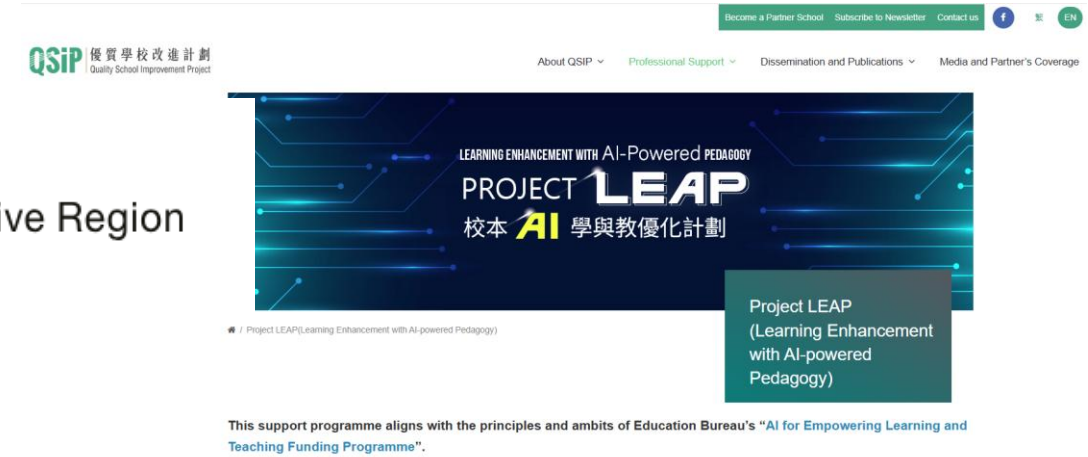
Language pedagogy genre, discourse, pragmatics, multilingual repertoire

AI affordances feedback, generation, adaptation, multimodal resources

Orchestration task sequence, evidence, teacher judgment, classroom norms

Ethics privacy, disclosure, bias, academic integrity, age-appropriate use

Education Bureau
The Government of the Hong Kong Special Administrative Region of the People's Republic of China



https://www.edb.gov.hk/en/edu-system/primary-secondary/applicable-to-primary-secondary/it-in-edu/pdp-ited.html
https://qsip.cuhk.edu.hk/en/project/leap/

Wrap-up 4: Equity and ethics

Responsible use of AI must protect data, learning, language identity, and human agency.

Access

Do all learners have meaningful, safe access?

Privacy

Are tools age-appropriate and data-safe?

Bias

Does AI generate bias-free output?

Validity

Can AI feedback/scoring handle local accents and diverse proficiency?

Attachment

Does social AI support, rather than replace, human care?

Human-centered AI in Hong Kong means designing for inclusion, multilingual identity, and contestable human judgment. Ethics should be built into task design.

Wrap-up 5: From pilots to school-based design

Policy funding may support change only when schools turn it into curriculum and evidence.

EDB AI funding programme

HK$500,000	one-off funding for successful publicly-funded primary and secondary schools
3+ subjects	AI-assisted teaching across at least three subjects/curriculum areas
6+ resources	teaching examples or sets of learning resources
3+ open classes	classroom demonstrations and sharing sessions

A school-based roadmap for language panels

1 term	**Explore**	A learning problem; 2–3 bounded AI tasks; collect student work
1 year	**Design**	Task bank; AI-use norms; lesson study; AI literacy routines
2–3 years	**Evaluate & scale**	Language outcomes; teacher workload; equity; cross-panel sharing

The unit of change is not the tool. It is the school-based pedagogical design.

中大推AI教育計劃助家長理解AI教育
料惠及逾1.1萬家長

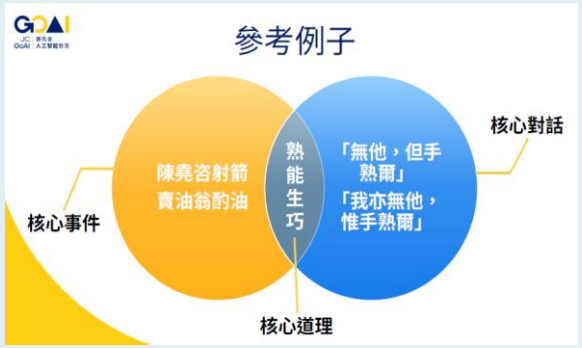
CUHK × JC GoAI: An AI-Assisted S1 Chinese lesson

Teaching **借事說理** (illustrating a principle through a story) through **賣油翁** . Students stay the analyst, and AI is the **文言文** tutor they question.

賣油翁 (Teaching slides)

S1 Chinese 中國語文中一課程

借事說理 | 熟能生巧



A Human–AI–Human Flow (Teacher Guide)

1. **動手玩 Go Play**
   Hook with “mastery” video clips and an oil-pouring mini-game.
   工具：瀝油神技小遊戲 AI 課文導賞影片

2. **動腦筋 Go Think**
   Students read 賣油翁 and circle the words they don’t know. In pairs they ask.
   工具：文言引導小助手（驗證推斷、照顧能力差異，而非直接給答案）

3. **齊創作 Go Make**
   Groups build a 事理圖.
   工具：人工智能寫作手法分析助手

4. **齊反思 Go Reflect**
   Groups share their maps.
   結尾提問聚焦「人工智能有哪些地方需要我們來補足的？」

AI supports 文言文理解 and 事理推理, but learners remain the interpreters and evaluators of language, culture, and meaning.

https://www.jcgoai.hk/en/our-programme/framework-pedagogy

Closing: Hong Kong schools can lead, not just adopt.

1 Evidence matters

AI can help language learning, but effects vary by learner, task, and contextual condition.

Three reviews

2 Design matters

Pedagogy-first, school-based inquiry, and professional sharing

Project LEAP

3 Agency matters most

System capacity, teacher-co-creation, equal access, ethics, and AI + subjects

JC GoAI

The future is AI-literate human communication: evidence-informed, design-based, ethically grounded, and agency-fostering.

Thank you!

Questions

alancheung@cuhk.edu.hk