

ซีรีส์ พลังทวี AI 100x



Hermes

AI AGENT

พลังทวี 100x

อัปเดต July 2026



เอเจนต์ที่อยู่บนเซิร์ฟเวอร์ของคุณเอง • ทำงาน 24 ชม.
จำคุณได้ • เก่งขึ้นทุกวัน • เลือกสมองได้เอง

Connect

Remember

Schedule

Delegate

Search



\$ hermes

คำนำ — ฉบับอัปเดต July 2026

สามเล่มก่อนในซีรีส์ "พลังทวี AI 100x" พาคุณรู้จัก ChatGPT, Gemini และ Claude — แชนบอทเก่งๆ ที่คุณเปิดเว็บขึ้นมาถามแล้วมันตอบ เล่มนี้ต่างออกไป และอาจเป็นก้าวที่น่าตื่นเต้นที่สุด

Hermes Agent ไม่ใช่แชนบอทบนเว็บของใคร แต่คือ **เอเจนต์ AI ที่คุณติดตั้งลงบนเซิร์ฟเวอร์ของคุณเอง** แล้วมันก็อยู่ตรงนั้น — ทำงาน 24 ชั่วโมง จำทุกอย่างที่คุณเคยบอก เก่งขึ้นเรื่อยๆ ยิ่งใช้ยิ่งรู้จักคุณ และที่สำคัญ **มันลงมือทำงานจริง** ไม่ใช่แค่ตอบข้อความ — รันคำสั่ง จัดการไฟล์ ค้นเว็บ ตอบลูกค้าในแชท และตั้งเวลาทำงานให้เองตอนคุณหลับ

Hermes สร้างโดย **Nous Research** เป็น **โอเพ่นซอร์ส (ฟรี)** ออกต้นปี 2026 แล้วโตพรวดจนกลายเป็นเอเจนต์ที่คนทั่วโลกใช้มากที่สุด จุดเด่นคือคุณ **เป็นเจ้าของมันเอง 100%** — ข้อมูลอยู่กับคุณ ปรับแต่งได้ทุกอย่าง และ **เลือกได้เองว่าจะให้มันใช้ "สมอง" ของโมเดลตัวไหน** (Claude, GPT, DeepSeek ฯลฯ) จับคู่ถูก-เก่งตามงานและงบ

เล่มนี้ยัด **6 โมดูลหลักของ Hermes — Connect • Remember • Schedule • Delegate • Search •**

Experiment เป็นส่วนหลัง และเขียนเพื่อ **คนทำงานและเจ้าของธุรกิจ ไม่ใช่โปรแกรมเมอร์** เราจะพาติดตั้งทีละขั้นแบบไม่ต้องกลัวโค้ด อัดแน่นด้วย **คำสั่งและ prompt ที่ก๊อปปี้ไปใช้ได้ทันทีเป็นร้อยชุด** ตัวอย่างจริงแบบไทยๆ และการบ้านสั้นๆ ให้ลงมือได้ใน 15 นาทีทุกบท

ปลายทางของเล่มนี้ไม่ใช่แค่ "รู้จัก AI อีกตัว" แต่คือการมี **ผู้ช่วยส่วนตัวที่เป็นของคุณจริงๆ** ทำงานให้คุณทั้งวันทั้งคืน

ขอให้สนุกกับการทวีพลังของคุณครับ

— ทีมงาน AI100x · กรกฎาคม 2026

สารบัญ

ส่วนที่ 1 · ทำไมปี 2026 คุณควรมี “เอเจนต์ส่วนตัว”	5
บทที่ 1 · Hermes Agent คืออะไร และทำไมมันเปลี่ยนเกม	6
ส่วนที่ 2 · ติดตั้งและเริ่มใช้งานจริง	17
บทที่ 2 · เตรียมของ: เซิร์ฟเวอร์ สมอง และค่าใช้จ่ายจริง	18
บทที่ 3 · ติดตั้ง Hermes ใน 20 นาที (ฉบับไม่ต้องเขียนโค้ด)	31
ส่วนที่ 3 · ศิลปะการสั่งงานเอเจนต์	42
บทที่ 4 · สั่งงานเอเจนต์ต่างจากสั่งแชทบอทยังไง	43
บทที่ 5 · คลังคำสั่ง 70+ สูตรใช้ได้ทันที	55
ส่วนที่ 4 · Connect: เอเจนต์เดียว อยู่ทุกแอปแชท	73
บทที่ 6 · เชื่อมเอเจนต์เข้า Telegram / Discord / แอปแชท	74
บทที่ 7 · ผู้ช่วยส่วนตัวในกระเป๋า: ใช้เอเจนต์ผ่านแชทจริง	85
ส่วนที่ 5 · Remember: เอเจนต์ที่จำคุณได้	95
บทที่ 8 · ระบบความจำ: ทำไมยิ่งใช้ยิ่งฉลาด	96
บทที่ 9 · Skills: เอเจนต์สร้างทักษะของตัวเอง	107
ส่วนที่ 6 · Schedule: ให้เอเจนต์ทำงานเองตอนคุณหลับ	118
บทที่ 10 · ตั้งงานอัตโนมัติด้วยภาษาคน	119
บทที่ 11 · เวอร์กโฟลว์อัตโนมัติสำหรับธุรกิจ	133
ส่วนที่ 7 · Search & Delegate: มือไม้และกำลังเสริม	144
บทที่ 12 · เครื่องมือครบมือ: ค้นเว็บ เบราวเซอร์ มองภาพ สร้างรูป พุดได้	145

ส่วนที่ 8 • เชื่อมเอเจนต์เข้ากับโลกจริง 165

บทที่ 14 • ต่อ MCP: ให้เอเจนต์เห็นข้อมูลธุรกิจคุณ 166

บทที่ 15 • เอเจนต์ทำการตลาดและปิดการขายให้คุณ 177

บทที่ 16 • บริการลูกค้า 24 ชม. ด้วยเอเจนต์บนแชท 189

ส่วนที่ 9 • โฮสต์ ปรับแต่ง และทำเงินจริง 200

บทที่ 17 • โฮสต์เอเจนต์ให้รับ 24/7 อย่างสบายใจ 201

บทที่ 18 • ปั่นบุคลิกและทำเอเจนต์ให้เป็น "แบรนด์" ของคุณ 211

บทที่ 19 • เปลี่ยน Hermes เป็นรายได้จริง 222

บทที่ 20 • เลือกสมองให้คุ้ม: เร็ว-ถูก-เก่ง 233

ส่วนที่ 10 • ปิดดัก และก้าวต่อไป 243

บทที่ 21 • ความปลอดภัยของเอเจนต์ที่ทำงานแทนคุณ 244

บทที่ 22 • แผน 30 วัน และอนาคตของเอเจนต์ส่วนตัว 255

ส่วนที่ 1

ทำไมปี 2026 คุณควรมี “เอเจนต์ส่วนตัว”

“Hermes ไม่ใช่เซทบอท — แต่คือเอเจนต์ที่อยู่บนเซิร์ฟเวอร์ของคุณเอง ทำงาน 24 ชม.
จำคุณได้ เก่งขึ้นทุกวัน”

บทที่ 1 — Hermes Agent คืออะไร และทำไมมันเปลี่ยนเกม

ลองนึกภาพว่าคุณมี "ผู้ช่วยส่วนตัว" คนหนึ่ง — คนที่อยู่ประจำ 24 ชั่วโมง ไม่หลับไม่พัก จำได้ทุกอย่างที่คุณเคยบอก รู้จักธุรกิจคุณ รู้จักนิสัยคุณ และยิ่งทำงานด้วยกันนานยิ่งเข้าใจคุณมากขึ้น คุณสั่งงานมันจากไลน์ตอนนั่งรถสั่งจากคอมตอนอยู่ร้าน หรือตั้งเวลาให้มันทำงานเองตอนคุณหลับก็ได้

นั่นแหละคือ **Hermes Agent** — และมันไม่ใช่แชทบอทแบบที่คุณเคยใช้

แชทบอท กับ เอเจนต์ ต่างกันตรงไหน

คุณคงเคยใช้ ChatGPT, Gemini หรือ Claude มาบ้าง สิ่งเหล่านี้คือ "แชทบอท" — เว็บถาม-ตอบของบริษัทเขา คุณเปิดเว็บ พิมพ์คำถาม มันตอบ จบแล้วก็จบเลย พงู่นี้เปิดใหม่ มันจำคุณไม่ได้ (เว้นแต่คุณตั้งค่าไว้บางส่วน) และที่สำคัญ — มันแค่ "พูด" ไม่ได้ "ลงมือทำ" จริง คุณให้มันไปกดส่งอีเมลจริง เปิดเว็บจริง หรือจัดการไฟล์ในเครื่องคุณไม่ได้

เอเจนต์ (agent) ต่างออกไป มันคือโปรแกรม AI ที่ **ติดตั้งอยู่บนเครื่อง/เซิร์ฟเวอร์ของคุณเอง** ทำงานตลอด 24 ชั่วโมง จำคุณได้ถาวร และ **ลงมือทำงานจริง** — รันคำสั่ง อ่าน-เขียนไฟล์ ค้นเว็บ ตั้งเวลาทำงานเอง คุยข้ามแอป แชทได้ทุกที่

พูดง่ายๆ: แชทบอทคือ "ที่ปรึกษา" ที่คุณต้องไปหาเขาถึงออฟฟิศทุกครั้ง — เอเจนต์คือ "พนักงาน" ที่นั่งอยู่ในเครื่องคุณ พร้อมทำงานให้ตลอดเวลา

ตารางเทียบชัดๆ: แชทบอท vs Hermes Agent

เพื่อให้เห็นภาพก่อนไปต่อ นี่คือตารางสรุปความต่างที่สำคัญที่สุด 7 ข้อ:

หัวข้อ	แชทบอท (ChatGPT/Gemini/Claude)	Hermes Agent
ข้อมูลอยู่ที่ไหน	เซิร์ฟเวอร์ของบริษัทเจ้าของแชทบอท	เซิร์ฟเวอร์ของคุณเอง
ความจำ	จำได้จำกัด/ตั้งค่าได้บางส่วน	จำถาวรข้ามเซสชัน สร้างสเกิลเองได้
ลงมือทำงานจริง	ทำไม่ได้ (ตอบอย่างเดียว)	ทำได้ — รันคำสั่ง อ่าน/เขียนไฟล์ ควบคุมเบราว์เซอร์
เชื่อมต่อแอปแชทอื่น	ส่วนใหญ่อยู่แค่เว็บ/แอปของตัวเอง	อยู่ได้ 20+ แพลตฟอร์ม ความจำเดียวกันหมด
ตั้งเวลาทำงานเอง	ทำไม่ได้	ทำได้ด้วยภาษาคน (natural-language cron)
เลือกสมอง (โมเดล)	ผูกกับเจ้าของแพลตฟอร์มเจ้าเดียว	สลับได้ตามงาน/งบ ผ่าน <code>/model</code>
ค่าใช้จ่าย	ค่าสมาชิกรายเดือนคงที่	จ่ายตามใช้จริง (VPS + โมเดล)

Hermes สร้างโดย **Nous Research** ทีมวิจัย AI สายโอเพ่นซอร์ส คำโปรยของมันคือ *"The Agent That Grows With You"* — เอเจนต์ที่เติบโตไปกับคุณ ตัวโปรแกรมเป็น **โอเพ่นซอร์ส (MIT License)** โหลดมาใช้ **ฟรี** สิ่งที่ต้องจ่ายมีแค่ (1) ที่รันเซิร์ฟเวอร์ และ (2) ค่าโมเดล AI ตามที่ใช้ (เดี๋ยวเจาะเรื่องเงินในบทที่ 2)

Hermes ออกครั้งแรกช่วง **กุมภาพันธ์ 2026** แล้วโตเร็วแบบก้าวกระโดด — บน GitHub (เว็บเก็บโค้ดของนักพัฒนาทั่วโลก) มีคนกดดาวให้ทะลุหลักแสนภายในไม่กี่เดือน และกลายเป็นเอเจนต์ที่คนใช้มากที่สุดตัวหนึ่งในปีนี้ เวอร์ชันล่าสุดกลางปี 2026 คือ **v0.18.0 "The Judgment Release"** (ยังเป็นเวอร์ชัน 0.x = พัฒนาเร็ว อัปเดตบ่อย ฟีเจอร์เพิ่มตลอด)

ลองสั่งงานดูก่อน — นี่คือน้ำตาการคุยกับเอเจนต์

ก่อนจะลงลึก ขอให้เห็นภาพก่อนว่า "การสั่งงานเอเจนต์" น้ำตาเป็นยังไง มันไม่ต้องเขียนโค้ด คุณแค่ **พิมพ์คุยเป็นภาษาคน** เหมือนสั่งลูกน้อง ตัวอย่างข้อความที่พิมพ์สั่ง Hermes ได้เลย:

สรุปอีเมลที่เข้ามาวันนี้ให้หน่อย เอาเฉพาะเรื่องที่ต้องตอบด่วน
แล้วร่างคำตอบเตรียมไว้ให้ด้วย เดี่ยวเราเช็กก่อนส่ง

เปิดเว็บคู่แข่งร้านนี้ [วางลิงก์] ดูว่าตอนนี้เขาขายอะไร ราคาเท่าไร
แล้วทำเป็นตารางเทียบกับสินค้าเราให้หน่อย

ทุกเช้า 8 โมง สรุปยอดขายเมื่อวานจากไฟล์ในเครื่อง
ส่งเข้าไลน์กลุ่มร้านให้เลย ไม่ต้องรอเราสั่ง

สังเกตว่าทั้งสามอันคือ "การพูดกับเอเจนต์" ธรรมดา ไม่มีศัพท์เทคนิค ไม่มีโค้ด แต่สิ่งที่มันทำต่างจากแชทบอทตรงที่ — มัน **ไปทำจริง** ทั้งเปิดอีเมลจริง เปิดเว็บจริง และตั้งเวลาทำงานเองได้จริง

Hermes ใช้ได้กับธุรกิจแบบไหนบ้าง

หลายคนอ่านมาถึงตรงนี้แล้วยังนึกภาพไม่ออกว่า "ธุรกิจแบบฉัน" จะเอาไปใช้ยังไง ลองดูตัวอย่างจากธุรกิจไทยหลายแบบ — ทุกอันด้านล่างคือสิ่งที่ Hermes ทำได้จริงด้วยความสามารถที่มีอยู่แล้ว:

ประเภทธุรกิจ	งานที่ยกให้เอเจนต์ทำ
ร้านอาหาร/เดลิเวอรี	สรุปยอดขาย ตอบลูกค้าคำถามเดิม เติมนราคาวัตถุดิบขึ้น
ร้านเสื้อผ้า/ของออนไลน์	เขียนแคปชั่น ตอบแชทลูกค้า เช็คสต็อกจากไฟล์
คลินิกความงาม/สปา	สรุปนัดหมายวันถัดไป ตอบ FAQ ราคาคอร์ส ส่งเตือนลูกค้า
ฟรีแลนซ์กราฟิก/นักเขียน	ร่างข้อเสนองาน ติดตามใบแจ้งหนี้ค้างจ่าย จัดคิวงาน
งานออฟฟิศ/ที่ปรึกษา	สรุปอีเมล นัดหมาย สรุปประชุมเป็นข้อความสั้น

ตัวอย่างคำสั่งที่ใช้ได้ในแต่ละธุรกิจ:

เรามีร้านเสื้อผ้าออนไลน์ขายผ่าน IG ชื่อ "@sabaidee.closet"
เขียนแคปชั่นขายเสื้อยืดคอลเลกชันใหม่ 3 แบบ โทนหวานๆ วัยรุ่นหน่อย

คลินิกเราชื่อ "Glow Aesthetic" มีคอร์สฉีดฟิลเลอร์กับเลเซอร์หน้าใส
มีลูกค้าทักมาถามราคาบ่อยๆ ช่วยร่างข้อความตอบมาตรฐานให้ 1 ชุด
เพื่อเวลาแอดมินไม่ว่างจะได้ส่งตอบได้เลย

เราเป็นฟรีแลนซ์กราฟิกดีไซน์ ช่วยเช็คไฟล์ invoice.xlsx ในเครื่อง
ว่ามีใบแจ้งหนี้ไหนเกิน 15 วันแล้วยังไม่จ่ายบ้าง สรุปรายชื่อมาให้

สรุปอีเมลที่ค้างตอบเกิน 2 วันให้หน่อย เรียงจากสำคัญไปน้อย
แล้วช่วยร่างนัดประชุมทีมสัปดาห์นี้ให้ 1 ฉบับ

5 จุดต่างที่ทำให้ Hermes เปลี่ยนเกม

นี่คือหัวใจว่าทำไม "เอเจนต์ส่วนตัว" ถึงต่างจากเซทบอทที่คุณคุ้นเคย

1. เป็นเจ้าของเอง 100%

Hermes รันบนเซิร์ฟเวอร์ของคุณ ข้อมูลอยู่กับคุณ ไม่ได้ฝากไว้บนคลาวด์ของบริษัทอื่น แปลว่า:

- บริษัทไหนจะปิดบัญชีคุณ เปลี่ยนกติกา ขึ้นราคา หรือลบประวัติแชทคุณทิ้งไม่ได้ เพราะมันเป็นของคุณ
- คุณปรับแต่งได้ทุกอย่าง จะตั้งชื่อมัน ตั้งบุคลิก ตั้งภาษา ให้ทำอะไรบ้าง คุณเองหมด
- ข้อมูลธุรกิจลับๆ ที่ไม่อยากให้ใครเห็น ก็อยู่ในเครื่องคุณ ไม่ได้ลอยไปไหน

จำไว้นะ: ธุรกิจเราซื้อร้าน "ครัวคุณย่าย" ขายอาหารอีสานส่งเดลิเวอรี่
ลูกค้าหลักคือออฟฟิศแถวอโศก เมนูขายดีคือส้มตำปูปลาร้ากับคอหมูย่าง
เวลาช่วยเราทำคอนเทนต์ ให้ยืดหยุ่น

เคล็ดลับ: ยิ่งคุณ "ป้อนข้อมูลธุรกิจ" ให้เอเจนต์จำไว้ตั้งแต่ต้น (ชื่อร้าน สินค้า ลูกค้า โทคนที่ชอบ) มันยิ่งตอบ
ได้ตรงใจเร็วขึ้น ไม่ต้องมานั่งอธิบายซ้ำทุกครั้งที่คุณสั่งงานใหม่

2. อยู่ทุกแอป ความจำเดียว

Hermes มีสิ่งที่เรียกว่า **gateway (ตัวกลางเชื่อมเอเจนต์เข้ากับแอปแชท)** ทำให้เอเจนต์ตัวเดียวของคุณโฟกัสได้
ในทุกแอปที่คุณใช้ — Telegram, Discord, Slack, WhatsApp, อีเมล และอีกกว่า 20 แพลตฟอร์ม (รวมถึง
Signal, SMS, Matrix, Microsoft Teams, Google Chat และ Home Assistant) หลักการคือ *"one agent, one
memory, every surface"* (เอเจนต์ตัวเดียว ความจำก้อนเดียว อยู่ได้ทุกหน้าจอ)

คุณคุยกับมันจาก Telegram ตอนเช้า พอเปิดคอมตอนบ่ายมันก็ยังจำเรื่องเมื่อเช้าได้ เพราะมันคือ "ตัวเดียวกัน"
ไม่ใช่คนละแอปคนละความจำ

เมื่อเช้าที่เราคุยกันเรื่องแคมเปญลดราคาสงกรานต์
สรุปไอเดียที่ตกลงกันไว้ให้หน่อย เดี่ยวจะเอาไปทำต่อ

เมื่อที่เราคุยกับมันผ่านมือถือตอนอยู่ในรถ
ตอนนี้เปิดคอมที่ร้านแล้ว ช่วยต่อยอดไอเดียเมื่อที่เป็นโพสต์ยาวให้หน่อย

เกร็ด: gateway ต่อได้มากกว่า 20 แพลตฟอร์มพร้อมกัน แปลว่าทีมงานหลายคนคุยกับเอเจนต์ตัวเดียวกัน
คนละแอปได้ แต่ทุกคนเห็น/ใช้ความจำก้อนเดียวกันหมด

3. ยิ่งใช้ยิ่งฉลาด

แชทบอทเริ่มจากศูนย์ทุกครั้ง แต่ Hermes มี **ความจำถาวร (persistent memory)** — มันจดสิ่งสำคัญเก็บไว้
เอง เรียนรู้โปรเจกต์คุณ จานิสัยการทำงานของคุณ และยังสร้าง "สเกิล (ทักษะ)" ของตัวเองขึ้นมาจากงานที่เคย
ทำ ยิ่งทำงานด้วยกันนาน มันยิ่งรู้จักคุณ ไม่ต้องเล่าบริบทซ้ำทุกครั้ง

ความจำของ Hermes ทำงานผ่านระบบสร้างโมเดล "ตัวตนผู้ใช้" ที่ชื่อ **Honcho** ผสมกับการค้นพบสนทนาเก่า
แบบเต็มข้อความ (FTS5) แล้วสรุปด้วย LLM ให้อัตโนมัติ — พูดย่อๆ คือมันไม่ได้แค่ "จำคำพูด" แต่พยายาม
"เข้าใจตัวตนคุณ" ไปด้วย

ที่เราชอบให้เขียนแคปชั่นแบบไหน จำได้ไหม
เขียนโพสต์ขายเมนูใหม่ให้หน่อย ใช้สไตล์เดิมที่เราชอบ

ช่วยค้นย้อนหลังที่ว่าเราเคยคุยเรื่องราคาส่งของกับซัพพลายเออร์เจ้าไหนไว้บ้าง
สรุปเป็นรายชื่อพร้อมราคาที่คุยกันได้

เคล็ดลับ: สกิลที่เอเจนต์สร้างเองจะถูกเก็บไว้ที่ `~/.hermes/skills/` และเรียกใช้ซ้ำได้ด้วยคำสั่ง `/skills` หรือพิมพ์ `<ชื่อสกิล>` ตรงๆ เลย — ไม่ต้องสอนมันใหม่ทุกครั้งทำงานคล้ายเดิม

4. ทำงานเองตอนคุณหลับ

Hermes ตั้งเวลาสั่งงานได้ **ด้วยภาษาคน** (ไม่ต้องรู้เรื่องเทคนิคการตั้งเวลาแบบเดิมๆ) แค่ว่า "ทุกเช้า" "ทุกวันจันทร์" "ทุกสิ้นเดือน" มันก็ทำงานให้เองตามนั้น แล้วส่งผลเข้าแชทให้คุณ

ทุกวันจันทร์ตอน 9 โมง ค้นข่าวเทรนด์อาหารในไทยสัปดาห์ที่ผ่านมา
สรุปเป็น 5 หัวข้อสั้นๆ ส่งเข้าไลน์ให้เรา ไว้หาไอเดียเมนูใหม่

ทุกคืนตอนเที่ยงคืน สืบกรองไฟล์ข้อมูลลูกค้ากับยอดขายเก็บไว้
แล้วบอกเราด้วยว่าเรียบร้อยไหม

ทุกวันที่ 25 ของเดือน เช็คไฟล์ `invoice.xlsx` ว่ามีใบแจ้งหนี้ไหน
ยังไม่ได้จ่าย แล้วช่วยร่างข้อความทวงเงินสุภาพๆ ส่งให้เราดูก่อน

เคล็ดลับ: งานอัตโนมัติๆ ที่ใช้โมเดลแพง อาจทำให้บิลบานได้ ให้พิมพ์ `/usage` เช็คการใช้งาน/ค่าใช้จ่ายเป็นระยะ (รายละเอียดเรื่องเงินเจาะลึกในบทที่ 2)

5. เลือกสมอง (โมเดล) ได้

นี่คือจุดที่หลายคนไม่เคยรู้ — Hermes **ไม่ผูกกับ AI เจ้าเดียว** ตัว Hermes คือ "ร่าง" ส่วน "สมอง" ที่ขับเคลื่อนมันคือ **โมเดล (model)** ที่คุณเลือกเสียบเข้าไปได้ จะใช้สมองของ Claude, GPT, DeepSeek หรือตัวอื่นก็ได้ สลับได้ตามงานและงบ

เปรียบเหมือนรถที่คุณเปลี่ยนเครื่องยนต์ได้ — งานเบาๆ ใช้เครื่องประหยัดน้ำมัน (โมเดลถูก) งานคิดหนัก ค่อยสลับเป็นเครื่องแรง (โมเดลพรีเมียม) ไม่ต้องจ่ายแพงตลอดเวลา

เปลี่ยนโมเดลง่ายมาก แค่พิมพ์คำสั่งสั้นๆ:

```
/model
```

หรือระบุตัวที่ต้องการเลย:

```
/model anthropic:claude-sonnet
```

งานที่ไม่ต้องคิดเยอะ อยากประหยัดสุด ก็สลับไปโมเดลถูกได้เลย:

```
/model deepseek:v4-flash
```

ตารางเทียบราคาโมเดล (กลางปี 2026, ตัวเลขโดยประมาณ)

ราคาต่อ 1 ล้าน token เปลี่ยนบ่อย ให้ยึดหลักการเรื่อง "ระดับราคา" เป็นหลัก แล้วเช็กราคาจริงอีกทีตอนใช้งาน:

ระดับ	โมเดลตัวอย่าง	เข้า / ออก (ต่อล้าน token)
ประหยัด	DeepSeek V4 Flash	~\$0.14 / \$0.28
ประหยัด	GPT-5.4 Nano	~\$0.20 / \$1.25
กลาง	Claude Haiku 4.5	~\$1 / \$5
พรีเมียม	Claude Sonnet 4.6	~\$3 / \$15
พรีเมียม	Claude Opus 4.8	~\$5 / \$25

ข้อควรระวัง: Nous เองก็มีโมเดลชื่อ **Hermes 4** (Hermes-4-70B / 405B) ให้เลือกใน Portal ราคาถูก แต่ **ไม่แนะนำให้ใช้ในลูปเอเจนต์** เพราะจูนมาเพื่อแชท/ให้เหตุผล ไม่ใช่เรียกเครื่องมือรัวๆ แบบที่เอเจนต์ต้องใช้งานจริง — งานที่ต้องให้เอเจนต์ "ลงมือทำ" ให้เลือกโมเดลสายเรียกเครื่องมือเก่ง เช่น Claude/GPT/DeepSeek แทน

Nous Portal คืออะไร

ถ้าไม่อยากไปสมัครกับผู้ให้บริการโมเดลทีละเจ้า Hermes มีทางเลือกคือ **Nous Portal** — จ่ายบิลเดียว เข้าถึงโมเดลได้ **300+ ตัว** พร้อมเครื่องมือหลัก 4 อย่าง (ค้นเว็บ, สร้างรูป, แปลงเสียง, เบราว์เซอร์บนคลาวด์) ในที่เดียว มีให้เลือก 4 แพคเกจ:

แพ็คเกจ	ค่าใช้จ่าย/เดือน	เครดิตที่ได้
Free	\$0	ตามโควตาฟรี
Plus	~\$20	~\$22
Super	~\$100	~\$110
Ultra	~\$200	~\$220

ทุกแพ็คเกจใช้ได้กับโมเดล 300+ ตัว + เครื่องมือหลัก 4 ตัว (มีเครดิตทยอยอด + โบนัสตอนสมัคร/ต่ออายุ) — เทียบให้เห็นภาพ: ค่ารันเอเจนต์ส่วนตัวขั้นต่ำใกล้เคียงค่าเซิร์ฟเวอร์รายเดือนทั่วไป (~\$20) แต่ได้ตัวที่ทำงานเอง 24 ชม. และเป็นเจ้าของเอง (รายละเอียดเรื่องเงินเต็มๆ อยู่ในบทที่ 2)

6 โมดูลหลัก — แผนที่ของทั้งเล่มนี้

ความสามารถของ Hermes จัดเป็น 6 กลุ่มใหญ่ ซึ่งเป็น "สันหลัง" ของหนังสือเล่มนี้ทั้งเล่ม เห็นภาพรวมไว้ก่อน แล้วเราจะเจาะทีละอันในบทถัดๆ ไป

โมดูล	ทำอะไร	ตัวอย่างสั้นๆ
Connect (เชื่อม)	เอเจนต์ตัวเดียวอยู่ได้ทุกแอปแชท 20+ แพลตฟอร์ม	คุยจาก Telegram มือถือได้ทั้งวัน
Remember (จำ)	ความจำถาวร จำโปรเจกต์+ตัวตนคุณ สร้างสเกลเอง	ไม่ต้องเล่าบริบทซ้ำ
Schedule (ตั้งเวลา)	สั่งงานอัตโนมัติด้วยภาษาคน	รายงานเช้าส่งเข้าแชทเอง
Delegate (มอบหมาย)	แตกเอเจนต์ลูก (subagent) ทำงานหลายอย่างขนานกัน	ทำคอนเทนต์หลายแพลตฟอร์มพร้อมกัน
Search (ค้น/มือไม้)	ค้นเว็บ เปิดเบราว์เซอร์ มองรูป สร้างรูป พุดได้	อ่านสลิปโอนเงิน ทำภาพโพสต์
Experiment (รันได้หลายที่)	เลือกที่รันได้ 6 แบบ บางแบบ "ลับ" ตอนไม่ใช้ประหยัด	รันบน VPS ถูกๆ 24 ชม.

อย่าเพิ่งกังวลถ้ายังไม่เข้าใจทุกช่อง — ทั้งเล่มนี้จะพาคุณไปที่ละโมดูล พร้อมตัวอย่างคำสั่งจริงที่ก๊อปปี้ไปใช้ได้ทันที ลองดูตัวอย่างการใช้อีก 2 โมดูลที่ยังไม่เห็นตัวอย่างข้างบน:

แต่งงานนี้เป็น subagent 3 ตัวพร้อมกัน: ตัวแรกเขียนแคปชั่น Instagram

ตัวที่สองเขียนแคปชั่น Facebook ตัวที่สามเขียนแคปชั่น TikTok

ทุกตัวขายเมนูใหม่ "ข้าวเหนียวมะม่วง" แล้วรวมผลมาให้เลือกทีเดียว

อ่านรูปสลิปโอนเงินที่แนบมานี้ เช็คว่ายอดตรงกับออเดอร์ #1023 ใหม่ (350 บาท)

ถ้าตรง ให้ทำเครื่องหมายในไฟล์ orders.csv ว่าจ่ายแล้ว

Delegate (subagent) ทำให้งานที่ต้องคิดหลายสายพร้อมกันเสร็จในที่เดียว ไม่ต้องรอทีละงาน ส่วน Search/มือไม่ทำให้เอเจนต์ "มองเห็น" รูปภาพและ "ลงมือ" ค้น-เปิดเว็บได้จริง ไม่ใช่แค่คุยเฉยๆ

หน้าตาการใช้งานจริง

Hermes ใช้ได้หลายทาง ไม่ได้มีแต่หน้าจอดำๆ น่ากลัวอย่างที่กลัว:

- **Hermes Desktop** — แอปเดสก์ท็อปหน้าต่างเหมือนแอปทั่วไป กดคลิกได้ ไม่ต้องแตะเทอร์มินัล แสดงผลเครื่องมือแบบสตรีมเรียลไทม์ให้เห็นว่าเอเจนต์กำลังทำอะไรอยู่
- **TUI ในเทอร์มินัล** — พิมพ์คุยได้หลายบรรทัด มีตัวช่วยเติมคำสั่ง เก็บประวัติแชท จัดจังหวะสั่งใหม่กลางคันได้
- **ผ่านแอปแชท** — Telegram, Discord ฯลฯ สั่งจากมือถือได้เหมือนแชทกับเพื่อน

และมันรันได้บนแทบทุกเครื่อง — macOS 12 ขึ้นไป, Windows 10-11, Linux ทุกดิสโทร แม้แต่มือถือ Android (ผ่านแอปชื่อ Termux) ก็รันได้ ตั้งแต่ VPS ถูกๆ ราวเดือนละ 5 ดอลลาร์ ไปจนถึงเครื่องแรงๆ

เกร็ดเล็กน้อย: ก่อนหน้านี้เอเจนต์ที่คนใช้กันชื่อ **OpenClaw** พอ Hermes มา คนย้ายกันเยอะ จนมีคำสั่งช่วยย้ายข้อมูลให้เลย (hermes claw migrate) ถ้าคุณเคยใช้ OpenClaw ก็ไม่ต้องเริ่มใหม่หมด แม้แต่สกิลเก่าก็นำเข้ามาใช้ต่อได้

ใครควรอ่านเล่มนี้

หนังสือเล่มนี้เขียนให้ **คนทำงาน เจ้าของธุรกิจ และฟรีแลนซ์ไทย ที่ไม่ใช่โปรแกรมเมอร์** โดยเฉพาะ ถ้าคุณเป็นแบบนี้ เล่มนี้ตรงกับคุณ:

- **เจ้าของร้าน/ธุรกิจออนไลน์** — อยากมีผู้ช่วยตอบลูกค้า ทำคอนเทนต์ สรุปรายอด มอนิเตอร์คู่แข่ง โดยไม่ต้องจ้างคนเพิ่ม
- **ฟรีแลนซ์** — อยากทำงานเร็วขึ้น รับงานได้มากขึ้น มีมือไม้ช่วยงานซ้ำๆ
- **คนทำงานประจำ** — อยากมีผู้ช่วยจัดการอีเมล นัดหมาย สรุปรประชุม งานเอกสาร
- **คนที่ลองใช้ AI มาบ้าง** — แต่เบื่อกับที่ต้องเล่าบริบทซ้ำทุกครั้ง และอยากได้ตัวที่ "ทำงานจริง" ไม่ใช่แค่ตอบ

ต้องเขียนโค้ดเป็นไหม? ไม่ต้อง เล่นนี้จะพาไปที่ละขั้นแบบไม่ต้องเขียนโค้ด สิ่งเดียวที่ Hermes ต่างจากแพลตฟอร์มคือมันต้องติดตั้งบนเซิร์ฟเวอร์และตั้งค่านิดหน่อยตอนแรก (เราสอนละเอียดในบทที่ 2-3) ถ้าติดจริงๆ จ้างคนช่วยตั้งครั้งเดียวก็ได้ หลังจากนั้นก็แค่ "พิมพ์คุย" เหมือนแชทปกติ

คำถามที่พบบ่อยก่อนเริ่ม (FAQ)

ไม่มีเซิร์ฟเวอร์เลย เริ่มได้ไหม? ได้ VPS (เครื่องเซิร์ฟเวอร์เช่ารายเดือน) ราคาถูกสุดก็เริ่มต้นได้แล้ว ต่ำสุดราวเดือนละ 5 ดอลลาร์ เราสอนวิธีเลือกและติดตั้งในบทที่ 2-3

ไม่เก่งคอมพิวเตอร์เลย จะติดตั้งเองไหวไหม? ขั้นตอนติดตั้งเป็นการรันคำสั่งสำเร็จรูปครั้งเดียว (ก๊อป-วาง) ไม่ต้องเขียนโค้ด ถ้าติดจริงๆ จ้างคนช่วยตั้งครั้งเดียวจบก็ได้ หลังจากนั้นใช้งานแค่ "พิมพ์คุย"

ถ้าเน็ตที่บ้าน/ร้านหลุด เอเจนต์จะหยุดทำงานไหม? ไม่หยุด เพราะ Hermes รันอยู่บน VPS ที่เปิดตลอด ไม่ใช่เครื่องที่บ้านคุณ ตราบใดที่ VPS ยังทำงาน เอเจนต์ก็ทำงานต่อได้แม้คุณจะปิดมือถือ/คอมไปแล้ว

ถ้าเลิกใช้ Hermes ข้อมูลจะหายไหม? ไม่หาย เพราะข้อมูลทั้งหมดอยู่ในเซิร์ฟเวอร์ของคุณเอง ไม่ใช่ของบริษัทไหน จะสำรอง ย้าย หรือลบเมื่อไหร่ก็ได้ตามใจคุณ

หนึ่งวันกับเอเจนต์ส่วนตัว

ลองจินตนาการวันทำงานของเจ้าของร้านออนไลน์คนหนึ่งที่มี Hermes:

7:00 ตื่นมาเปิดไลน์ เจอข้อความจากเอเจนต์: "สรุปยอดขายเมื่อวาน 47 ออเดอร์ ยอดรวม 18,600 บาท เมนูขายดีสุดคือส้มตำปูปลาร้า มีลูกค้าถามเรื่องจัดส่งต่างจังหวัด 3 ราย ร่างคำตอบเตรียมไว้แล้วนะ"

10:30 ระหว่างเตรียมของ พิมพ์สั่งจากมือถือ: "ทำแคปชั่นโปรฯ ส่งฟรีเมนูใหม่ 3 เวอร์ชัน" — สักครูได้มาเลย เลือกอันที่ชอบ กดโพสต์

14:00 มีลูกค้าถามคำถามเดิมๆ ในเพจ เอเจนต์ที่เชื่อมต่อไว้ตอบ FAQ ให้อัตโนมัติ เรื่องไหนตอบไม่ได้ค่อยส่งต่อให้คุณ

22:00 ก่อนนอน สั่ง: "พรุ่งนี้เข้าเช็คราคาวัตถุดิบจากร้านส่งเจ้าประจำ ถ้าขึ้นเกิน 10% เตือนเราด้วย" — แล้วปิดเครื่องเข้านอน เอเจนต์ยังทำงานต่อบนเซิร์ฟเวอร์

นี่ไม่ใช่นิยาย ทุกอย่างในภาพนี้ Hermes ทำได้จริงด้วยความสามารถที่มีในตัว และเราจะสอนคุณทำทุกอันในเล่มนี้ ตัวอย่างคำสั่งที่ใช้ก็เช่น:

สรุปยอดขายเมื่อวานจากไฟล์ orders.csv ในเครื่อง

บอกจำนวนออเดอร์ ยอดรวม และเมนูขายดี 3 อันดับแรก

เขียนแคปชั่นขายเมนูใหม่ "ลาบหมูทอด" 3 เวอร์ชัน

เวอร์ชันแรกเน้นความอร่อย อันที่สองเน้นโปรฯ ส่งฟรี อันที่สามสั้นกระชับ

พรุ่งนี้ 7 โมงเช้า เช็คราคาน้ำมันพืชกับหมูจากเว็บร้านส่ง [ลิงก์]

ถ้าแพงขึ้นเกิน 10% จากอาทิตย์ก่อน แจ้งเตือนเราทันที

อีกหนึ่งวัน: ฟรีแลนซ์กราฟิกดีไซน์เนอร์

ไม่ใช่แค่ร้านค้าที่ใช้ Hermes ได้ ฟรีแลนซ์ก็ใช้ประโยชน์ได้เต็มๆ ลองดูวันทำงานของนักออกแบบกราฟิกอิสระคนหนึ่ง:

8:30 เปิดคอมเจอสรุปจากเอเจนต์: "มีใบแจ้งหนี้ 2 ใบเกิน 15 วันยังไม่จ่าย ลูกค้า A กับ C ร่างข้อความทวงสุภาพๆ ไว้ให้แล้ว รอคุณกดส่ง"

11:00 พิมพ์ส่ง: "ช่วยร่างข้อเสนองานออกแบบโลโก้ให้ลูกค้าใหม่ ราคาตามเรตปกติของเรา" — ได้ร่างมาปรับนิดหน่อยก็ส่งได้เลย

17:00 ก่อนเลิกงาน ส่งเอเจนต์ให้ทำสรุปงานที่ทำวันนี้ ส่งเข้า Discord ทีมงาน

ช่วยร่างข้อเสนองานออกแบบโลโก้ให้ลูกค้าใหม่ชื่อ "ร้านกาแฟดอยหลวง"

ใช้เรตราคาปกติของเรา (จำได้จากที่เคยบอกไว้) ส่งเป็นข้อความสุภาพ

สรุปงานที่ทำเสร็จวันนี้เป็นข้อความสั้นๆ 3-4 บรรทัด

ส่งเข้า Discord ช่องทีมงานให้หน่อย

สิ่งที่ต้องพูดตรงๆ

เล่มนี้จะไม่ขายฝัน ขอบอกตามจริง 3 ข้อ:

1. **Hermes เทคนิคกว่าแซทบอททั่วไปนิดหน่อย** — ตอนติดตั้งครั้งแรกต้องตั้งค่าบ้าง แต่เราพาไปที่ละขั้น และทำครั้งเดียวจบ
2. **มันยังเป็นเวอร์ชัน 0.x** — พัฒนาเร็ว เมนู/คำสั่งอาจขยับได้ ให้ยึด "หลักการ" ที่เล่มนี้สอน แล้วเช็คเอกสารทางการประกอบ
3. **เอเจนต์ที่ทำงานแทนได้จริง = ต้องคุมความปลอดภัยจริงจัง** โดยเฉพาะเรื่องเงิน กฎหมาย และการลบข้อมูล ต้องเปิดระบบขออนุมัติก่อนลงมือ และตรวจงานเสมอ (เราเจาะเรื่องนี้เต็มๆ ในบทที่ 21)

เคล็ดลับความปลอดภัยเบื้องต้น: Hermes มีระบบ **ขออนุมัติคำสั่ง (command approval)** ให้เปิดใช้เสมอสำหรับงานสำคัญ มีการกำหนดสิทธิ์การใช้เครื่องมือ (authorization) และแยกสภาพแวดล้อมทำงานด้วยคอนเทนเนอร์ (container isolation) เพื่อกันงานหนึ่งไปกระทบอีกงานหนึ่ง — อย่าปิดระบบขออนุมัติทิ้งเพียงเพราะอยากให้เร็วขึ้น โดยเฉพาะงานที่แตะเงินหรือข้อมูลลูกค้า

พร้อมแล้ว บทต่อไปเราจะเตรียมของ — ดูว่าต้องมีอะไรบ้าง และค่าใช้จ่ายจริงเท่าไร ก่อนลงมือติดตั้งจริงในบทที่ 3

⚡ ลงมือใน 15 นาที

ยังไม่ต้องติดตั้งอะไร แค่ "เตรียมสมอง" ให้พร้อม — หยิบกระดาษหรือเปิดโน้ตในมือถือ แล้วเขียน 3 อย่างนี้:

1. **งานซ้ำๆ ที่คุณทำทุกวัน/ทุกสัปดาห์** เขียนมาให้ได้ 5 อย่าง (เช่น สรุปรายชื่อลูกค้าเก่าๆ เช็คราคาคูแ่ง) — นี่คือการงานแรกที่จะโยนให้เอเจนต์ทำ
2. **ข้อมูลธุรกิจ 3–5 บรรทัด** ที่คุณอยากให้เอเจนต์จำ (ชื่อร้าน ขายอะไร ลูกค้าคือใคร โทคนที่ชอบ) — เก็บไว้ ตอนตั้งค่าเสร็จเราจะป้อนให้มันจำ
3. เข้าไปดูหน้าเว็บทางการของ Hermes ที่ hermes-agent.nousresearch.com เพื่อดูภาพรวมและความเคลื่อนไหวล่าสุด

เท่านี้คุณก็พร้อมสำหรับบทที่ 2 แล้ว

📌 สรุปบทนี้

- **Hermes Agent = เอเจนต์ AI ที่รันบนเซิร์ฟเวอร์ของคุณเอง** ทำงาน 24 ชม. จำคุณได้ และลงมือทำงานจริง — ต่างจากแชทบอท (ChatGPT/Gemini/Claude) ที่เป็นแค่เว็บถาม-ตอบของบริษัทเขา
- **5 จุดต่างที่เปลี่ยนเกม:** เป็นเจ้าของเอง 100% / อยู่ทุกแอปความจำเดียว / ยิ่งใช้ยิ่งฉลาด / ทำงานเองตอนคุณหลับ / เลือกสมอง (โมเดล) ได้
- ทั้งเล่มยัด **6 โมดูล** เป็นส้นหลัง: Connect · Remember · Schedule · Delegate · Search · Experiment
- Hermes เป็น **โอเพ่นซอร์สฟรี** ของ **Nous Research** ตัวโปรแกรมไม่เสียเงิน จ่ายแค่ค่าเซิร์ฟเวอร์ + ค่าโมเดล และเป็นเอเจนต์ที่โตเร็วสุดตัวหนึ่งในปี 2026
- ใช้ได้กับธุรกิจหลายแบบ ตั้งแต่ร้านอาหาร ร้านเสื้อผ้า คลินิก ไปจนถึงฟรีแลนซ์และงานออฟฟิศ — เลือกโมเดล/แพ็คเกจ Nous Portal ให้เหมาะงบตัวเองได้
- เล่มนี้เขียนให้ **เจ้าของธุรกิจ/ฟรีแลนซ์/คนไทยที่ไม่ต้องเขียนโค้ด** — สั่งงานด้วยการ "พิมพ์คุยเป็นภาษาคน" เหมือนสั่งลูกน้อง

ส่วนที่ 2

ติดตั้งและเริ่มใช้งานจริง

“ของฟรี รันบนเครื่องคุณเอง — จ่ายแค่ที่รันกับค่าสมอง แล้วเริ่มได้ใน 20 นาที”

บทที่ 2 — เตรียมของ: เซิร์ฟเวอร์ สมอง และค่าใช้จ่ายจริง

ก่อนจะลงมือติดตั้ง Hermes จริงในบทหน้า บทนี้พาคุณเช็คของให้ครบก่อน 2 อย่างเท่านั้น: "ที่รัน" (เครื่องที่ให้ Hermes ทำงานอยู่ตลอด 24 ชม.) กับ "สมอง" (โมเดล AI ที่ให้เอเจนต์ใช้คิด) แล้วจะรู้ทันทีว่าทั้งเดือนต้องจ่ายจริงเท่าไร ไม่มีค่าใช้จ่ายแอบแฝง

ง่ายๆ: **Hermes = ฟรี** (โอเพ่นซอร์ส MIT License) แต่ **"ที่รัน"** กับ **"สมอง"** ไม่ฟรี — สองอย่างนี้แหละคือค่าใช้จ่ายทั้งหมดที่คุณต้องจ่ายทุกเดือน ไม่มีค่าลิขสิทธิ์ตัวโปรแกรมซ่อนอยู่

บทนี้จะพาคุณไล่ทีละขั้น: เลือกที่รัน → เลือกสมอง → คำนวณต้นทุนจริงเป็นตัวเลข → ดูตัวอย่างธุรกิจไทยหลายแบบ เพื่อให้ปิดบทนี้แล้วรู้ทันทีว่าตัวเองจะจ่ายเดือนละเท่าไร ไม่ต้องเดา ไม่ต้องกลัวบิลซอกที่หลัง

ของ 2 อย่างที่ต้องมีก่อนเริ่ม

1. เครื่องรัน (ที่ให้ Hermes ทำงานอยู่) — เลือกได้ 3 แบบ:

- **VPS (Virtual Private Server** — เครื่องเซิร์ฟเวอร์เสมือน เช่ารายเดือน ไม่ต้องมีฮาร์ดแวร์เอง) ราคาเริ่มต้นราวๆ **\$5/เดือน** — เหมาะสุดถ้าอยากให้อเอเจนต์ทำงาน 24 ชม. จริง เช่น ตอบลูกค้าตอนดึก สรุยอดขายตอนตี 5 หรือมอนิเตอร์ราคาคู่แข่งทั้งคืน
- **คอมพิวเตอร์** — Mac / Windows / Linux เครื่องที่ใช้อยู่แล้ว ไม่มีค่าใช้จ่ายเพิ่มด้านเครื่อง แต่ต้องเปิดเครื่องทิ้งไว้ตลอดถ้าอยากใหทำงานต่อเนื่อง เหมาะกับช่วงทดลองเล่นก่อนตัดสินใจลงทุน
- **มือถือ Android ผ่าน Termux** (แอปจำลองเทอร์มินัลลินุกซ์บนมือถือ) — ลงได้จริง เหมาะกับทดลองเล่นระหว่างเดินทางหรือดูว่า Hermes หน้าตาเป็นยังไง แต่ไม่เหมาะเป็นเครื่องหลักที่ต้องรันทั้งวันทั้งคืน (แบตเตอรี่หมด หรือปิดแอปพลาตก็หยุดทำงานทันที)

2. สมอง (โมเดล AI) — ตัวที่ให้เอเจนต์คิด ตัดสินใจ เรียกเครื่องมือ คุณเลือกได้เองว่าจะใช้ค่ายไหน ราคาต่างกันหลายเท่าตัว ซึ่งเป็นหัวใจของบทนี้ และเป็นจุดที่ Hermes ต่างจากแชทบอทเจ้าเดียวอย่างชัดเจน — ไม่มีใครบังคับให้คุณผูกกับโมเดลใดโมเดลหนึ่งตลอดไป

ทั้งสองอย่างแยกจากกันเด็ดขาด: เครื่องรันคือ "บ้าน" ที่เอเจนต์อยู่ ส่วนโมเดลคือ "สมอง" ที่ใส่เข้าไปในบ้านนั้น จะย้ายบ้านหรือเปลี่ยนสมองเมื่อไหร่ก็ได้ ไม่ต้องเริ่มนับหนึ่งใหม่

เปรียบเทียบ 3 ที่รัน แบบเห็นภาพ

ก่อนตัดสินใจ ลองดูตารางเทียบข้อดี-ข้อจำกัดของแต่ละแบบให้ชัดๆ อีกที จะได้เลือกถูกตั้งแต่ครั้งแรก:

ที่รัน	ทำงาน 24 ชม. ได้จริงไหม	ค่าใช้จ่ายเพิ่ม	เหมาะกับใคร
VPS	ได้ (นี่คืองานหลักของมัน)	~\$5-25/เดือน	คนที่ต้องการให้เอเจนต์ทำงานจริงจัง ตอบลูกค้า/ตั้งเวลาอัตโนมัติ
คอมตัวเอง	ได้ ถ้าเปิดทิ้งไว้ตลอด	ไม่มี (แต่ต้องเปิดเครื่อง+เน็ตตลอด)	คนที่อยากทดลองเล่นก่อนตัดสินใจลงทุน
มือถือ Termux	ไม่แนะนำ (แบตเตอรี่/แอปถูกปิดง่าย)	ไม่มี	ทดลองดูหน้าตาระหว่างเดินทาง ไม่ใช่เครื่องหลัก

เคล็ดลับ: ถ้ายังไม่แน่ใจว่าจะใช้ Hermes จริงจังไหม ให้เริ่มจากคอมตัวเองหรือ Termux ทดลองใช้งานสัก 3-5 วันก่อน พอเห็นว่าคุ้มค่าจริง ค่อยย้ายไป VPS เพื่อให้ทำงาน 24 ชม. — ย้ายทีหลังได้ไม่เสียของเดิม เพราะความจำและสกินที่เอเจนต์เรียนรู้ไว้ ย้ายตามไปได้

ระบบที่รองรับ

อุปกรณ์	รองรับ
macOS	12 ขึ้นไป
Windows	10 – 11
Linux	ทุกดิสโทร
Android	ผ่าน Termux

ถ้าใช้คอมตัวเอง เช็คสเปกเครื่องคร่าวๆ ก่อนได้ด้วยคำสั่งนี้ (Linux/Mac):

```
free -h && nproc
```

คำสั่งนี้จะบอก RAM ที่มี กับจำนวนคอร์ CPU — Hermes รันได้ตั้งแต่เครื่องสเปกต่ำสุด **1 vCPU / 1 GB RAM** ไปจนถึงเครื่อง GPU แรงๆ ถ้าเครื่องคุณมีมากกว่านี้แปลว่าใช้ได้สบายแล้ว ไม่ต้องซื้อเครื่องใหม่

Hermes Desktop และ TUI — ถ้าไม่อยากพิมพ์คำสั่งตลอดเวลา

หลายคนได้ยินคำว่า "ติดตั้งบนเซิร์ฟเวอร์" แล้วกลัวว่าต้องนั่งพิมพ์คำสั่งต่างๆ ทั้งวัน ความจริงมีตัวช่วยลดความน่ากลัวลงมาก:

- **Hermes Desktop** — แอปเดสก์ท็อปหน้าต่างเหมือนแอปทั่วไป ไม่ต้องเปิดเทอร์มินัลก็ใช้งานได้ เห็นเครื่องมือที่เอเจนต์กำลังเรียกใช้แบบสตรีมเรียลไทม์ (เห็นสดๆ ว่ากำลังทำอะไรอยู่)

- **TUI (Text User Interface** — หน้าจอเทอร์มินัลที่จัดรูปแบบสวยงามใช้งานง่ายกว่าพิมพ์คำสั่งดิบๆ) พิมพ์ข้อความหลายบรรทัดได้ มี autocomplete คำสั่ง / ให้เอง ดูประวัติแชทย้อนหลังได้ และขัดจังหวะสั่งงานใหม่กลางคันได้โดยไม่ต้องรอมันทำงานเดิมจบก่อน

พูดง่ายๆ คือต่อให้ไม่ถนัดพิมพ์คำสั่งเลย ก็ยังมีหน้าต่างที่เป็นมิตรให้ใช้ได้ ไม่ต้องท่องคำสั่งทุกตัวก่อนเริ่ม

VPS เลือกยังไง ไม่ต้องคิดมาก

ถ้าไม่มีคอมที่เปิดทิ้งไว้ 24 ชม. ได้ให้เช่า VPS ราคาเริ่มต้นมีตั้งแต่ **~\$4-25/เดือน** ขึ้นกับสเปก ตัวอย่างผู้ให้บริการที่ราคาเข้าถึงง่ายอย่าง Hostinger เริ่มต้นราวๆ **\$6/เดือน** ก็เพียงพอสำหรับเริ่มต้นแล้ว สิ่งที่ต้องดูตอนเลือกมีแค่ 2 อย่าง:

เช็คว่าก่อนเช่า VPS:

1. RAM ขั้นต่ำ 1 GB (ยิ่งใช้หลายเครื่องมือพร้อมกัน ยิ่งอยากได้มากกว่านี้)
2. เลือกที่ตั้งเซิร์ฟเวอร์ (region) ใกล้ประเทศไทยหรือสิงคโปร์ เพื่อความเร็วในการเชื่อมต่อ

ไม่ต้องกังวลว่าต้องเลือกให้ถูกที่สุดตั้งแต่แรก — สเปกเล็กสุดก็เริ่มได้ ถ้าใช้ไปแล้วรู้สึกว่าการเลือกหลายตัวทำงานพร้อมกันแล้วหน่วง ค่อยอัปเกรด RAM ที่หลังได้ ผู้ให้บริการ VPS ส่วนใหญ่ให้อัปเกรดได้ทันทีโดยไม่ต้องย้ายเครื่องใหม่

เคล็ดลับ: อย่าเพื่อสเปกเกินจำเป็นตั้งแต่วันแรก เพราะ RAM/CPU ที่ซื้อเกินมาแต่ไม่ได้ใช้ คือเงินที่จ่ายทิ้งเปล่าทุกเดือน เริ่มจากสเปกเล็กสุด ใช้จริงสัก 1-2 สัปดาห์ค่อยดูว่าต้องอัปเกรดไหม

ตัวโปรแกรม Hermes ฟรี — จ่ายแค่ 2 อย่างเท่านั้น

Hermes Agent ตัวโปรแกรมเป็น **โอเพ่นซอร์ส MIT License** ดาวน์โฮลด์และติดตั้งได้ฟรี 100% ไม่มีค่าลิขสิทธิ์ ไม่มีค่าสมัครสมาชิกรายเดือนของตัวโปรแกรมเอง สิ่งที่ต้องจ่ายจริงมีแค่:

ค่าใช้จ่ายทั้งหมดของ Hermes = (1) ค่าที่รัน (VPS) + (2) ค่าสมอง (โมเดล AI ตามที่ใช้จริง)

พูดง่ายๆ คือเหมือนซื้อ "ตัวรถฟรี" แต่ต้องจ่ายค่าที่จอดรถ (VPS) กับค่าน้ำมัน (โมเดล AI ที่คิดเป็นการทำงานจริง ยิ่งสั่งงานเยอะ/โมเดลแพง ยิ่งจ่ายเยอะ) เจ้าของร้านหลายคนพอได้ยินว่า "ต้องมีเซิร์ฟเวอร์เอง" ก็กลัวว่าจะแพง แต่พอคำนวณจริงจะเห็นว่าค่าใช้จ่ายเริ่มต้นถูกกว่าจ้างพนักงานพาร์ทไทม์คนหนึ่งมาก และเอเจนต์ทำงานได้ทั้งวันทั้งคืนโดยไม่ต้องพักผ่อน

เลือกมองตามงบ — ตารางราคาต่อล้าน token

token คือหน่วยนับข้อความที่ AI ใช้อ่าน/เขียน (ประมาณคำหรือส่วนของคำ) ยิ่งคุ้ยยาว ยิ่งให้เอเจนต์อ่านไฟล์เยอะ ยิ่งใช้ token เยอะ ราคาโมเดลจะคิดเป็น "ต่อ 1 ล้าน token" แยกฝั่ง "เข้า" (สิ่งที่คุณป้อนให้อ่าน เช่น คำสั่ง เอกสาร ประวัติแชท) กับ "ออก" (สิ่งที่ AI ตอบกลับหรือลงมือทำ)

ระดับ	โมเดลตัวอย่าง	ราคา เข้า / ออก (ต่อล้าน token)	เหมาะกับ
ประหยัด	DeepSeek V4 Flash	~\$0.14 / \$0.28	งานเบา ทำซ้ำๆ เยอะ เช่น ตอบแชทลูกค้าทั่วไป จัดหมวดหมู่ข้อมูล
ประหยัด	GPT-5.4 Nano	~\$0.20 / \$1.25	งานสั้นกระชับ ไม่ต้องคิดซับซ้อน
กลาง	Claude Haiku 4.5	~\$1 / \$5	งานประจำวันที่ต้องแม่นยำขึ้นมาน้อย เช่น สรุปรายงาน ร่างข้อความ
พรีเมียม	Claude Sonnet 4.6	~\$3 / \$15	งานคิดวิเคราะห์ วางแผน เขียนโค้ด/เนื้อหาซับซ้อน
พรีเมียม	Claude Opus 4.8	~\$5 / \$25	งานสำคัญที่ต้องการความแม่นยำสูงสุด ตัดสินใจธุรกิจสำคัญ

ราคาข้างต้นเป็นตัวเลขราวๆ กลางปี 2026 — เปลี่ยนได้ตลอด ให้เช็กหน้าราคาจริงอีกทีก่อนตัดสินใจ

คิด token เป็นตัวเงินจริงยังงี้ ไม่ต้องเดา

หลายคนเห็นตัวเลข "ต่อล้าน token" แล้วนึกภาพไม่ออกว่าเดือนหนึ่งจะใช้ไปเท่าไร ลองดูตัวอย่างคำนวณคร่าวๆ นี้:

ร้านค้าออนไลน์ตอบแชทลูกค้าเฉลี่ยวันละ 200 ข้อความ

- แต่ละข้อความ (รวมคำถามลูกค้า+ประวัติแชท+คำตอบ) กินราว 300-500 token
- 200 ข้อความ/วัน x 400 token เฉลี่ย x 30 วัน = ~2.4 ล้าน token/เดือน
- ใช้ DeepSeek V4 Flash (~\$0.14 เข้า / \$0.28 ออก ต่อล้าน) → ตกราวๆ \$1-2/เดือน
- ถ้าใช้ Claude Sonnet 4.6 แทน (~\$3 เข้า / \$15 ออก ต่อล้าน) → ตกราวๆ \$20-35/เดือน แพงกว่ากันสิบเท่า

ตัวเลขนี้เป็นแค่การประมาณให้เห็นภาพ ตัวเลขจริงต่างกันตามความยาวบทสนทนาและงานที่สั่ง แต่หลักการสำคัญคือ งานที่ทำซ้ำๆ ทุกวัน ต่างราคากันนิดเดียวต่อครั้ง แต่พอกันด้วยจำนวนครั้งทั้งเดือน ผลต่าง

จะกลายเป็นหลักสิบหลักร้อยเหรียญทันที นี่คือเหตุผลที่ต้องเลือกสมองให้เหมาะกับงาน ไม่ใช่ใช้ตัวแพงสุดตลอด

งานแบบไหน ควรใช้สมองระดับไหน

ไม่ต้องใช้โมเดลแพงที่สุดตลอดเวลา — คนที่คุ้มค่าใช้จ่ายจะ "จับคู่งานกับสมอง" ให้เหมาะสม ลองดูตัวอย่างจริงตามธุรกิจ:

ร้านขายเสื้อผ้าออนไลน์

- ตอบแชทลูกค้าถามไซส์/สี/ราคา (เยอะ วันละหลายร้อยข้อความ) → ประหยัด
- วางแผนแคปชั่นแคมเปญเปิดตัวคอลเลกชันใหม่ → พรีเมียม

คลินิกความงาม

- นัดหมายคนไข้ ตอบคำถามทั่วไป → ประหยัด/กลาง
- ร่างสคริปต์ให้คำปรึกษาคนไข้เคสซับซ้อน → พรีเมียม

ฟรีแลนซ์กราฟิก

- จัดระเบียบไฟล์งาน สรุปีเมลลูกค้า → กลาง
- ช่วยคิดคอนเซปต์งานออกแบบให้ลูกค้ารายใหญ่ → พรีเมียม

ตัวอย่างต้นทุนอีก 4 ธุรกิจไทย

นอกจาก 3 ธุรกิจข้างบน ลองดูอีก 4 แบบที่พบบ่อยในไทย เพื่อให้เห็นว่าหลักการ "จับคู่งานกับสมอง" ใช้ได้กับแทบทุกอาชีพ:

ร้านอาหาร/ร้านกาแฟ (รับออเดอร์ผ่าน LINE OA)

- ตอบลูกค้าถามเมนู/เวลาเปิด-ปิด/โปรโมชั่น (เยอะมาก ทุกวัน) → ประหยัด
- คิดไอเดียเมนูใหม่/แคปชั่นโปรโมชั่นเทศกาล → กลาง/พรีเมียม

สำนักงานบัญชี/ฟรีแลนซ์ทำบัญชี

- สรุปรายรับ-รายจ่ายลูกค้าประจำเดือน จัดหมวดหมู่ใบเสร็จ → กลาง
- วิเคราะห์ภาษี/วางแผนโครงสร้างบัญชีเคสซับซ้อน → พรีเมียม

นายหน้าอสังหาริมทรัพย์

- ตอบคำถามทั่วไปเรื่องทำเล/ราคา/นัดดูบ้าน → ประหยัด/กลาง
- เขียนคำบรรยายทรัพย์สินแบบขายง่าย + วิเคราะห์เปรียบเทียบราคาดตลาด → พรีเมียม

ครูสอนพิเศษ/ติวเตอร์ออนไลน์

- ตอบคำถามนักเรียนเรื่องตารางเรียน/ค่าเรียน → ประหยัด
- ออกแบบแผนการสอนเฉพาะเด็กแต่ละคน/ข้อสอบเจาะจุดอ่อน → พรีเมียม

จะเห็นแพทเทิร์นซ้ำๆ ทุกธุรกิจ: **งานที่ตอบซ้ำ/ตอบเยอะ ใช้โมเดลประหยัด** **งานที่ต้องคิดวิเคราะห์ลึก/ผลกระทบสูง ใช้โมเดลพรีเมียม** ไม่มีธุรกิจไหนต้องใช้โมเดลแพงสุดตลอดทั้งวัน

สั่งเลือกสมองด้วยคำสั่งเดียว

ตอนตั้งค่าครั้งแรก เปิดเมนูเลือกโมเดลด้วยคำสั่ง:

```
hermes model
```

จะขึ้นรายชื่อผู้ให้บริการและโมเดลให้เลือก ระหว่างคุยงานอยู่ ถ้าอยากสลับโมเดลกลางคัน ใช้ slash command ในแชทได้ทันทีตามรูปแบบ `/model [provider:model]` เช่น

```
/model deepseek:deepseek-v4-flash  
/model anthropic:claude-haiku-4.5  
/model anthropic:claude-sonnet-4.6
```

ชื่อรุ่นย่อยที่แสดงจริงอาจเปลี่ยนไปตามที่ผู้ให้บริการอัปเดต ให้ยึดชื่อที่ขึ้นจริงตอนพิมพ์ hermes model เป็นหลัก อย่าจำตายตัวจากหนังสือเล่มนี้

ตัวอย่างการใช้งานจริง: ร้านขายเสื้อผ้าออนไลน์ให้เอเจนต์ตอบแชทลูกค้าทั้งวัน (งานเยอะ ไม่ซับซ้อน) → ตั้งเป็น DeepSeek V4 Flash ประหยัดสุด แต่พอถึงเวลาให้เอเจนต์ช่วยวางแผนแคมเปญโปรโมชันปลายเดือน (งานคิดหนัก) → สลับไป Claude Sonnet 4.6 ชั่วโมงด้วยคำสั่งเดียว ไม่ต้องตั้งค่าใหม่ทั้งระบบ พอวางแผนเสร็จก็สลับกลับโหมดประหยัดได้ทันที

เคล็ดลับ: ถ้าไม่แน่ใจว่างานหนึ่งควรใช้โมเดลระดับไหน ให้เริ่มจากตัวประหยัดที่สุดก่อนเสมอ ลองสั่งงานดูสักครั้ง ถ้าคำตอบยังคลุมเครือหรือพลาดจุดสำคัญ ค่อยสลับขึ้นไปโมเดลระดับสูงกว่า วิธีนี้ประหยัดกว่าการเปิดโมเดลแพงสุดไว้ตั้งแต่ต้นเสมอ

Nous Portal คืออะไร คุ่มใหม่

Nous Portal คือทางเลือก "จ่ายรวมทีเดียว" แทนการต่อตรงกับแต่ละค่ายโมเดลเอง สมัครครั้งเดียวได้เข้าถึง **300+ โมเดล** ในบิลเดียว พร้อม **เครื่องมือหลัก 4 ตัว** (ค้นเว็บผ่าน Firecrawl, สร้างรูปผ่าน FAL, แปลงข้อความเป็นเสียงผ่าน OpenAI TTS, เบราวเซอร์บนคลาวด์ผ่าน Browser Use) ตั้งค่าเชื่อมต่อได้ด้วยคำสั่งเดียวตอนติดตั้งครั้งแรก:

```
hermes setup --portal
```

คำสั่งนี้จะเปิดหน้าเว็บให้ล็อกอิน (OAuth — ระบบยืนยันตัวตนผ่านเว็บ ไม่ต้องพิมพ์รหัสผ่านซ้ำในเทอร์มินัล) เชื่อมบัญชี Portal เข้ากับ Hermes ของคุณ หลังเชื่อมต่อเสร็จ ทุกครั้งที่สั่งงานหรือสลับโมเดล บิลจะไปรวมอยู่ที่บัญชี Portal บัญชีเดียว ไม่ต้องผูกบัตรเครดิตหลายใบกับหลายเว็บไซต์

"เครดิต" ในที่นี้คือยอดเงินที่เติมไว้ล่วงหน้าในบัญชี Portal ใช้จ่ายกับโมเดลไหนก็ได้จนกว่าจะหมด คล้ายๆ บัตรเติมเงินโทรศัพท์ ไม่ใช่แพ็คเกจจำกัดจำนวนครั้งใช้งาน

แพ็คเกจ Nous Portal

แพ็คเกจ	ราคา/เดือน	เครดิตที่ได้	เหมาะกับใคร
Free	\$0	—	ทดลองเล่นก่อนตัดสินใจ
Plus	~\$20	~\$22	เจ้าของร้าน/ฟรีแลนซ์คนเดียว ใช้งานเบาถึงกลางทุกวัน
Super	~\$100	~\$110	ทีมเล็ก ใช้โมเดลฟรีเมียมบ่อย หลายงานพร้อมกัน
Ultra	~\$200	~\$220	ใช้งานหนักทั้งวัน หรือรันเอเจนต์ให้ลูกค้าหลายราย

เทียบให้เห็นภาพ: ค่ารันเอเจนต์ส่วนตัวขั้นต้นของคุณ ใกล้เคียงค่าเซทบอทรายเดือนทั่วไป (~\$20) แต่ได้ตัวที่ทำงานเอง 24 ชม. และเป็นเจ้าของเอง

ข้อดีของ Portal คือ **ไม่ต้องไปสมัครหลายเจ้า ไม่ต้องผูกบัตรหลายใบ** เห็นค่าใช้จ่ายรวมทีเดียว เหมาะกับคนที่อยากสลับโมเดลไปมาบ่อยๆ ตามงาน โดยไม่ต้องคำนวณแยกทีละเจ้า ส่วนถ้าใช้แค่โมเดลเดียวประจำและอยากคุมราคาต่อ token ให้ละเอียดที่สุด การต่อตรงกับผู้ให้บริการเอง (ผ่าน OpenRouter หรือ endpoint ของค่ายนั้นโดยตรง) ก็ยังทำได้เหมือนเดิม เลือกทางไหนไม่ผิด ขึ้นกับว่าอยากคุมบิลแบบไหนมากกว่ากัน

Portal หรือ ต่อตรงเอง — เปรียบเทียบให้ชัด

หัวข้อ	Nous Portal	ต่อตรงกับค่ายโมเดล/OpenRouter
จำนวนบิลที่ต้องดู	บิลเดียว รวมทุกโมเดล+เครื่องมือ	แยกบิลตามค่ายที่ใช้
เครื่องมือหลัก 4 ตัว (ค้นเว็บ/สร้างรูป/TTS/เบราวเซอร์)	รวมมาให้แล้วในแพ็คเกจ	ต้องไปสมัคร/จ่ายแยกเอง
ความยืดหยุ่นสลับโมเดล	สลับได้ทันทีในบิลเดียว	สลับได้เหมือนกัน แต่บิลแยกค่าย
เหมาะกับ	คนอยากง่าย ไม่อยากดูหลายบิล	คนอยากคุมราคาต่อ token ละเอียดสุด/ใช้โมเดลเดียวประจำ

เช็การใช้งานและบิลด้วย /usage

ไม่ว่าจะต่อตรงหรือผ่าน Portal ให้เช็คยอดใช้จ่ายเป็นระยะด้วย slash command:

```
/usage
```