แบบจำลองเชิงแนวคิดเพื่ออธิบายความเสี่ยงของ AI Hallucination

(A Conceptual Mathematical Model for Explaining the Risk of AI Hallucination)

รายงานทางวิชาการ

10 กันยายน พ.ศ. 2569

บทคัดย่อ

รายงานฉบับนี้นำเสนอ **แบบจำลองเชิงแนวคิด (Conceptual Model)** เพื่ออธิบายความเสี่ยงของการเกิดปรากฏการณ์ AI Hallucination โดยใช้แนวคิดทางคณิตศาสตร์ในรูปแบบของฟังก์ชันโลจิสติก (Logistic Function) เป็นเครื่องมือในการจำลองความสัมพันธ์เชิงมโนทัศน์ระหว่างปัจจัยต่าง ๆ ที่อาจส่งผลต่อความเสี่ยงดังกล่าว แบบจำลองนี้ **ไม่ใช่สูตรทางวิทยาศาสตร์ที่ผ่านการวัดหรือพิสูจน์จากข้อมูลจริงของระบบ AI ใด ๆ** หากแต่เป็นเครื่องมือเชิงแนวคิดที่สร้างขึ้นเพื่อช่วยให้เข้าใจว่า Hallucination เป็นผลลัพธ์ที่เกิดจากปัจจัยหลายประการทำงานร่วมกัน มิใช่ผลจากปัจจัยใดปัจจัยหนึ่งเพียงลำพัง

สารบัญ

1 บทนำ **4**

1.1 ความหมายของ AI Hallucination 4

1.2 เหตุใด Hallucination จึงไม่ได้เกิดจากปัจจัยเดียว 4

1.3 วัตถุประสงค์ของการสร้าง Mathematical Model 4

2 แนวคิดของแบบจำลอง **5**

2.1 คำนิยามของตัวแปร 5

2.2 ตารางสรุปตัวแปร 5

3 การกำหนดสมมติฐาน **6**

4 การสร้างสมการ Mathematical Model **6**

4.1 เหตุผลในการเลือกใช้ Logistic Function 6

4.2 สมการหลักของแบบจำลอง 6

4.3 ความหมายของสัมประสิทธิ์ 7

4.4 ตัวอย่างพารามิเตอร์สำหรับสร้างกราฟ 7

5 การลดแบบจำลองให้สามารถแสดงผลในกราฟ 2 มิติ **8**

5.1 ตารางสรุปสถานการณ์ทั้ง 4 กรณี 8

5.2 กรณีที่ 1: ความซับซ้อนต่ำ + Prompt คุณภาพสูง + บริบทเพียงพอ . . . 8

5.3 กรณีที่ 2: ความซับซ้อนสูง + Prompt คุณภาพสูง + บริบทเพียงพอ . . . 8

5.4 กรณีที่ 3: ความซับซ้อนต่ำ + Prompt คุณภาพต่ำ + บริบทไม่เพียงพอ . . 9

5.5 กรณีที่ 4: ความซับซ้อนสูง + Prompt คุณภาพต่ำ + บริบทไม่เพียงพอ . . 9

6 สมการสำหรับนำไปกรอกใน Desmos **9**

7 การวิเคราะห์กราฟ **10**

7.1 ความหมายของแกน 10

7.2 ความหมายของเส้นแต่ละเส้นและตำแหน่งที่แตกต่างกัน 10

7.3 จุดที่ความเสี่ยงเพิ่มขึ้นอย่างรวดเร็ว 10

7.4 ความสัมพันธ์ระหว่างปัจจัยต่าง ๆ 11

8 การกำหนดระดับความเสี่ยง **11**

9 การวิเคราะห์เชิงคณิตศาสตร์ **11**

9.1 พฤติกรรมของ Logistic Function 11

9.2 อนุพันธ์ของฟังก์ชันโลจิสติก 12

9.3 ผลของการเปลี่ยนค่าพารามิเตอร์ 12

10 การเชื่อมโยงกับการทดลอง AI **12**

11 ข้อจำกัดของแบบจำลอง **13**

12 สรุป **13**

1 บทนำ

1.1 ความหมายของ AI Hallucination

AI Hallucination หมายถึงปรากฏการณ์ที่แบบจำลองปัญญาประดิษฐ์เชิงรู้สร้าง (Generative AI) สร้างเนื้อหาที่ดูมีความสมเหตุสมผลในเชิงภาษา แต่ไม่ถูกต้องตามข้อเท็จจริง ไม่สอดคล้องกับแหล่งข้อมูลอ้างอิง หรือเป็นข้อมูลที่ไม่มีอยู่จริง ปรากฏการณ์นี้เกิดขึ้นได้แม้ในแบบจำลองที่มีประสิทธิภาพสูง เนื่องจากกลไกการทำงานของแบบจำลองภาษาเป็นการทำนายลำดับคำที่มีความน่าจะเป็นสูงสุด มิใช่การตรวจสอบความจริงจากฐานความรู้ที่ตายตัว

1.2 เหตุใด Hallucination จึงไม่ได้เกิดจากปัจจัยเดียว

จากลักษณะการทำงานของแบบจำลองภาษาขนาดใหญ่ (Large Language Model) ความเสี่ยงในการเกิด Hallucination มีที่มาจากหลายองค์ประกอบที่ทำงานร่วมกัน อาทิ

- ความไม่แน่นอนหรือความคลุมเครือของข้อมูลที่แบบจำลองมีอยู่ในตัว
- ความซับซ้อนของคำถามหรือภารกิจที่ผู้ใช้ร้องขอ
- คุณภาพและความชัดเจนของ Prompt ที่ผู้ใช้ป้อนเข้าไป
- ปริมาณและความเกี่ยวข้องของบริบทหรือข้อมูลสนับสนุนที่แบบจำลองได้รับ

ด้วยเหตุนี้ การอธิบายความเสี่ยงของ Hallucination ด้วยปัจจัยเดียวจึงไม่สามารถสะท้อนความซับซ้อนของปรากฏการณ์ได้อย่างครบถ้วน จำเป็นต้องอาศัยแบบจำลองที่สามารถรวมปัจจัยหลายมิติเข้าไว้ด้วยกัน

1.3 วัตถุประสงค์ของการสร้าง Mathematical Model

รายงานฉบับนี้มีวัตถุประสงค์เพื่อ

1. สร้างแบบจำลองเชิงแนวคิดที่แสดงความสัมพันธ์ระหว่างปัจจัยต่าง ๆ กับความเสี่ยงของ Hallucination ในรูปแบบคณิตศาสตร์
2. ใช้แบบจำลองดังกล่าวเป็นกรอบความคิดสำหรับการวิเคราะห์และเปรียบเทียบสถานการณ์ต่าง ๆ อย่างเป็นระบบ
3. เน้นย้ำว่าแบบจำลองนี้เป็นเครื่องมือเชิงแนวคิด (Conceptual Tool) มิใช่มาตรวัดทางวิทยาศาสตร์ที่ใช้ประเมิน Hallucination ของระบบ AI จริง

2 แนวคิดของแบบจำลอง

แบบจำลองในรายงานนี้กำหนดตัวแปรหลัก 5 ตัว โดยตัวแปรทั้งหมดถูกกำหนดให้อยู่ในช่วง $[0, 1]$ เพื่อความสะดวกในการคำนวณและการนำไปแสดงผลเป็นกราฟ

2.1 คำนิยามของตัวแปร

- U **(Data Uncertainty)** คือความไม่แน่นอนของข้อมูลที่แบบจำลองมีอยู่เกี่ยวกับหัวข้อที่ถูกถาม ค่า U สูงหมายถึงแบบจำลองมีข้อมูลภายในที่คลุมเครือ ขัดแย้งกันเอง หรือไม่เพียงพอเกี่ยวกับเรื่องนั้น ในขณะที่ค่า U ต่ำหมายถึงแบบจำลองมีความมั่นใจในข้อมูลที่เกี่ยวข้องกับหัวข้อนั้นสูง
- C **(Question Complexity)** คือความซับซ้อนของคำถามหรือภารกิจ ครอบคลุมทั้งความซับซ้อนเชิงโครงสร้างของคำถาม จำนวนขั้นตอนของการให้เหตุผลที่จำเป็น และความเฉพาะทางของเนื้อหา
- P **(Prompt Quality)** คือคุณภาพของ Prompt ที่ผู้ใช้ป้อนให้แบบจำลอง ครอบคลุมความชัดเจน ความเฉพาะเจาะจง และโครงสร้างของคำสั่ง ค่า P สูงหมายถึง Prompt มีความชัดเจนและมีข้อกำหนดที่ช่วยลดความกำกวม
- B **(Context)** คือปริมาณและความเกี่ยวข้องของบริบทหรือข้อมูลสนับสนุนที่แบบจำลองได้รับ เช่น เอกสารอ้างอิงที่แนบมา หรือข้อมูลจากระบบดึงข้อมูล (Retrieval) ค่า B สูงหมายถึงแบบจำลองมีข้อมูลสนับสนุนที่เพียงพอและตรงประเด็น
- H **(Hallucination Risk)** คือความเสี่ยงในการเกิด Hallucination ซึ่งเป็นผลลัพธ์ของแบบจำลอง มีค่าอยู่ในช่วง $[0, 1]$ โดย $H = 0$ หมายถึงความเสี่ยงต่ำที่สุด และ $H = 1$ หมายถึงความเสี่ยงสูงที่สุดตามแนวคิดของแบบจำลอง

2.2 ตารางสรุปตัวแปร

ตัวแปร	**ความหมาย**	**ช่วงค่า**	**ทิศทางผลต่อ H**
U	ความไม่แน่นอนของข้อมูล	$[0, 1]$	เพิ่ม H
C	ความซับซ้อนของคำถาม	$[0, 1]$	เพิ่ม H
P	คุณภาพของ Prompt	$[0, 1]$	ลด H
B	บริบท/ข้อมูลสนับสนุน	$[0, 1]$	ลด H
H	ความเสี่ยงของ Hallucination	$[0, 1]$	ตัวแปรผลลัพธ์

3 การกำหนดสมมติฐาน

แบบจำลองนี้ตั้งอยู่บน **สมมติฐานเชิงแนวคิด (Conceptual Assumptions)** ต่อไปนี้ ซึ่งมิใช่ข้อสรุปเชิงสถิติที่ได้จากการทดลองหรือการเก็บข้อมูลจริงจากระบบ AI ใด ๆ หากแต่เป็นสมมติฐานเชิงเหตุผลที่ใช้เป็นพื้นฐานในการสร้างแบบจำลองทางคณิตศาสตร์

1. เมื่อ U เพิ่มขึ้น (ข้อมูลมีความไม่แน่นอนมากขึ้น) ความเสี่ยง H มีแนวโน้มเพิ่มขึ้น
2. เมื่อ C เพิ่มขึ้น (คำถามมีความซับซ้อนมากขึ้น) ความเสี่ยง H มีแนวโน้มเพิ่มขึ้น
3. เมื่อ P เพิ่มขึ้น (Prompt มีคุณภาพดีขึ้น) ความเสี่ยง H มีแนวโน้มลดลง
4. เมื่อ B มีความเพียงพอและเกี่ยวข้องกับคำถามมากขึ้น ความเสี่ยง H มีแนวโน้มลดลง

ข้อควรระวัง: สมมติฐานข้างต้นเป็นการสร้างกรอบความคิดเชิงเหตุผลเพื่อใช้ประกอบการสร้างแบบจำลองทางคณิตศาสตร์เท่านั้น มิได้มีที่มาจากการวัดผลเชิงประจักษ์ (Empirical Measurement) จากระบบ AI จริงแต่อย่างใด

4 การสร้างสมการ Mathematical Model

4.1 เหตุผลในการเลือกใช้ Logistic Function

เนื่องจากค่าความเสี่ยง H ถูกกำหนดให้อยู่ในช่วง $[0, 1]$ เพื่อสื่อความหมายในเชิงสัดส่วนหรือความน่าจะเป็นเชิงแนวคิด ฟังก์ชันโลจิสติก (Logistic Function) จึงเป็นฟังก์ชันที่เหมาะสมสำหรับการจำลองลักษณะเช่นนี้ เนื่องด้วยคุณสมบัติดังนี้

- ให้ค่าผลลัพธ์อยู่ในช่วง $(0, 1)$ เสมอ ไม่ว่าตัวแปรต้นจะมีค่าเท่าใด
- มีลักษณะเป็นเส้นโค้งรูปตัว S (Sigmoid) ซึ่งสะท้อนแนวคิดที่ว่าความเสี่ยงจะเพิ่มขึ้นอย่างช้า ๆ ในช่วงแรก เพิ่มขึ้นอย่างรวดเร็วในช่วงกลาง แล้วค่อย ๆ เข้าใกล้ขีดจำกัดบนในช่วงท้าย
- สามารถรวมตัวแปรหลายตัวเข้าด้วยกันในรูปแบบเชิงเส้นผ่านตัวแปรกลาง Z ก่อนแปลงเป็นค่าความเสี่ยง

4.2 สมการหลักของแบบจำลอง

กำหนดให้ความเสี่ยงของ Hallucination เป็นฟังก์ชันของตัวแปรทั้งสี่ ดังสมการ

$$H(U, C, P, B) = \frac{1}{1 + e^{-Z}} \tag{1}$$

โดยที่ Z คือตัวแปรเชิงเส้นรวม (Linear Combination) ซึ่งนิยามดังนี้

$$Z = aU + bC - cP - dB + k \tag{2}$$

4.3 ความหมายของสัมประสิทธิ์

- a คือสัมประสิทธิ์ที่แสดงน้ำหนักผลกระทบของความไม่แน่นอนของข้อมูล U ต่อความเสี่ยง โดยมีเครื่องหมายบวก สะท้อนว่า U ยิ่งสูง ความเสี่ยงยิ่งเพิ่ม
- b คือสัมประสิทธิ์ที่แสดงน้ำหนักผลกระทบของความซับซ้อนของคำถาม C โดยมีเครื่องหมายบวกเช่นเดียวกัน
- c คือสัมประสิทธิ์ที่แสดงน้ำหนักผลกระทบของคุณภาพ Prompt P โดยมีเครื่องหมายลบ สะท้อนว่า P ยิ่งสูง ความเสี่ยงยิ่งลดลง
- d คือสัมประสิทธิ์ที่แสดงน้ำหนักผลกระทบของบริบท B โดยมีเครื่องหมายลบเช่นเดียวกับ c
- k คือค่าคงที่ปรับฐาน (Baseline Constant) ที่ใช้เลื่อนตำแหน่งของเส้นโค้งบนแกน Z เพื่อกำหนดจุดสมดุลเริ่มต้นของแบบจำลองในเชิงแนวคิด

4.4 ตัวอย่างพารามิเตอร์สำหรับสร้างกราฟ

เพื่อความชัดเจนในการนำแบบจำลองไปแสดงผลเป็นกราฟ รายงานนี้กำหนด **ตัวอย่างค่าพารามิเตอร์เชิงสมมติ** ดังนี้

$$a = 8, \quad b = 4, \quad c = 3, \quad d = 2, \quad k = -5 \tag{3}$$

ค่าดังกล่าวเป็นเพียงตัวเลขสมมติที่กำหนดขึ้นเพื่อสาธิตพฤติกรรมของแบบจำลอง **มิใช่ค่าที่ได้จากการฝึกฝนหรือประมาณค่าจากข้อมูลจริงของระบบ AI ใด ๆ** เหตุผลของการเลือกขนาดสัมประสิทธิ์มีดังนี้

- $a = 8$ มีค่าสูงที่สุด สะท้อนแนวคิดว่าความไม่แน่นอนของข้อมูลเป็นปัจจัยที่มีอิทธิพลสูงต่อความเสี่ยง
- $b = 4$ รองลงมา สะท้อนว่าความซับซ้อนของคำถามมีอิทธิพลปานกลางถึงสูง
- $c = 3$ และ $d = 2$ เป็นสัมประสิทธิ์เชิงลบ สะท้อนบทบาทของ Prompt และบริบทในการบรรเทาความเสี่ยง โดย Prompt มีน้ำหนักการบรรเทาที่มากกว่าบริบทเล็กน้อยในแบบจำลองนี้
- $k = -5$ เป็นค่าคงที่ที่เลื่อนจุดสมดุลของเส้นโค้งให้อยู่ในตำแหน่งที่เหมาะสมสำหรับการสาธิตในช่วง $U \in [0, 1]$

5 การลดแบบจำลองให้สามารถแสดงผลในกราฟ 2 มิติ

เนื่องจากเครื่องมือกราฟสองมิติ เช่น Desmos ไม่สามารถแสดงฟังก์ชันที่มีตัวแปรอิสระหลายตัวพร้อมกันได้อย่างสะดวก รายงานนี้จึงกำหนดให้

$$x = U \tag{4}$$

เป็นตัวแปรอิสระบนแกนนอน และตรึงค่า C, P, B ให้เป็นค่าคงที่ในแต่ละสถานการณ์ เพื่อให้สามารถแสดงความสัมพันธ์ระหว่าง H และ U ได้ในลักษณะกราฟเส้นสองมิติ โดยกำหนดสถานการณ์เชิงแนวคิดทั้งหมด 4 กรณี ดังนี้

5.1 ตารางสรุปสถานการณ์ทั้ง 4 กรณี

กรณี	**ลักษณะสถานการณ์**	C	P	B
1	ความซับซ้อนต่ำ + Prompt คุณภาพสูง + บริบทเพียงพอ	0.2	0.9	0.9
2	ความซับซ้อนสูง + Prompt คุณภาพสูง + บริบทเพียงพอ	0.9	0.9	0.9
3	ความซับซ้อนต่ำ + Prompt คุณภาพต่ำ + บริบทไม่เพียงพอ	0.2	0.2	0.2
4	ความซับซ้อนสูง + Prompt คุณภาพต่ำ + บริบทไม่เพียงพอ	0.9	0.2	0.2

5.2 กรณีที่ 1: ความซับซ้อนต่ำ + Prompt คุณภาพสูง + บริบทเพียงพอ

แทนค่า $C = 0.2$, $P = 0.9$, $B = 0.9$ ลงในสมการ (2) โดยใช้พารามิเตอร์จากสมการ (3)

$$Z_1 = 8x + 4(0.2) - 3(0.9) - 2(0.9) - 5 = 8x - 8.7 \tag{5}$$

$$H_1(x) = \frac{1}{1 + e^{-(8x-8.7)}} \tag{6}$$

กรณีนี้เป็นสถานการณ์ที่เอื้ออำนวยต่อการลดความเสี่ยงมากที่สุดในบรรดา 4 กรณี เนื่องจากค่าคงที่ในสมการ Z_1 ติดลบมาก ทำให้ต้องอาศัยค่า U ที่สูงมากจึงจะทำให้ความเสี่ยง H เข้าใกล้ค่าสูง

5.3 กรณีที่ 2: ความซับซ้อนสูง + Prompt คุณภาพสูง + บริบทเพียงพอ

แทนค่า $C = 0.9$, $P = 0.9$, $B = 0.9$

$$Z_2 = 8x + 4(0.9) - 3(0.9) - 2(0.9) - 5 = 8x - 5.9 \tag{7}$$

$$H_2(x) = \frac{1}{1 + e^{-(8x-5.9)}} \tag{8}$$

แม้ Prompt จะมีคุณภาพสูงและบริบทเพียงพอเช่นเดียวกับกรณีที่ 1 แต่ความซับซ้อนของคำถามที่สูงขึ้นทำให้ค่าคงที่ในสมการ Z_2 ติดลบน้อยลง ส่งผลให้เส้นโค้งของกรณีนี้เลื่อนไปทางซ้ายเมื่อเทียบกับกรณีที่ 1 นั่นคือมีความเสี่ยงสูงขึ้นที่ค่า U เดียวกัน

5.4 กรณีที่ 3: ความซับซ้อนต่ำ + Prompt คุณภาพต่ำ + บริบทไม่เพียงพอ

แทนค่า $C = 0.2$, $P = 0.2$, $B = 0.2$

$$Z_3 = 8x + 4(0.2) - 3(0.2) - 2(0.2) - 5 = 8x - 5.2 \tag{9}$$

$$H_3(x) = \frac{1}{1 + e^{-(8x-5.2)}} \tag{10}$$

แม้คำถามจะไม่ซับซ้อน แต่คุณภาพ Prompt ที่ต่ำและบริบทที่ไม่เพียงพอทำให้เส้นโค้งของกรณีนี้เลื่อนไปทางซ้ายมากกว่ากรณีที่ 1 และ 2 แสดงให้เห็นว่าปัจจัยด้าน Prompt และบริบทมีบทบาทสำคัญไม่น้อยไปกว่าความซับซ้อนของคำถาม

5.5 กรณีที่ 4: ความซับซ้อนสูง + Prompt คุณภาพต่ำ + บริบทไม่เพียงพอ

แทนค่า $C = 0.9$, $P = 0.2$, $B = 0.2$

$$Z_4 = 8x + 4(0.9) - 3(0.2) - 2(0.2) - 5 = 8x - 2.4 \tag{11}$$

$$H_4(x) = \frac{1}{1 + e^{-(8x-2.4)}} \tag{12}$$

กรณีนี้เป็นสถานการณ์ที่มีความเสี่ยงสูงที่สุดในบรรดาทั้ง 4 กรณี เนื่องจากปัจจัยทุกตัวล้วนส่งผลในทิศทางที่เพิ่มความเสี่ยงพร้อมกัน ทำให้เส้นโค้งเลื่อนไปทางซ้ายมากที่สุด นั่นคือแม้ค่า U จะยังไม่สูงมาก ความเสี่ยง H ก็สามารถเข้าใกล้ค่าสูงได้อย่างรวดเร็ว

6 สมการสำหรับนำไปกรอกใน Desmos

ผู้ใช้สามารถนำสมการต่อไปนี้ไปกรอกในโปรแกรม Desmos (`www.desmos.com/calculator`) เพื่อสร้างกราฟของทั้ง 4 กรณีบนแกนเดียวกัน โดยแกนนอนแทนค่า $x = U$ และแกนตั้งแทนค่า H

กรณี	สมการสำหรับกรอกใน Desmos
1	`y = 1/(1+e^(-(8x-8.7)))`
2	`y = 1/(1+e^(-(8x-5.9)))`
3	`y = 1/(1+e^(-(8x-5.2)))`
4	`y = 1/(1+e^(-(8x-2.4)))`

แนะนำให้กำหนดขอบเขตการแสดงผลใน Desmos เป็น $x \in [0, 1]$ และ $y \in [0, 1]$ เพื่อให้สอดคล้องกับนิยามของตัวแปรในแบบจำลอง

7 การวิเคราะห์กราฟ

7.1 ความหมายของแกน

- **แกน X** แทนค่า U หรือความไม่แน่นอนของข้อมูล มีค่าอยู่ในช่วง $[0, 1]$
- **แกน Y** แทนค่า H หรือความเสี่ยงของการเกิด Hallucination มีค่าอยู่ในช่วง $[0, 1]$

7.2 ความหมายของเส้นแต่ละเส้นและตำแหน่งที่แตกต่างกัน

เส้นโค้งทั้ง 4 เส้นล้วนมีรูปทรงแบบ S (Sigmoid) เหมือนกัน แต่มีตำแหน่งบนแกน x แตกต่างกัน โดยเส้นของกรณีที่ 4 อยู่ค่อนไปทางซ้ายสุด (เพิ่มขึ้นเร็วที่สุด) ตามด้วยกรณีที่ 3 กรณีที่ 2 และกรณีที่ 1 อยู่ทางขวาสุด (เพิ่มขึ้นช้าที่สุด) ตามลำดับ ความแตกต่างของตำแหน่งนี้สะท้อนว่า เมื่อปัจจัยเสี่ยง (ความซับซ้อนสูง คุณภาพ Prompt ต่ำ บริบทไม่เพียงพอ) สะสมมากขึ้น เส้นโค้งจะเลื่อนไปทางซ้าย นั่นคือแบบจำลองจะถึงระดับความเสี่ยงสูงได้ที่ค่า U ที่ต่ำกว่า

7.3 จุดที่ความเสี่ยงเพิ่มขึ้นอย่างรวดเร็ว

บริเวณที่เส้นโค้งมีความชันสูงที่สุดคือบริเวณใกล้จุด $Z = 0$ หรือ $H = 0.5$ ซึ่งเป็นจุดที่ความเสี่ยงเพิ่มขึ้นอย่างรวดเร็วที่สุดเมื่อ U เปลี่ยนแปลงเพียงเล็กน้อย ตำแหน่งของจุดนี้บนแกน x สามารถคำนวณได้จากเงื่อนไข $Z = 0$ ดังตัวอย่างในกรณีที่ 4 ซึ่งจุดดังกล่าวอยู่ที่ $x = 2.4/8 = 0.3$ แสดงว่าภายใต้เงื่อนไขความซับซ้อนสูง Prompt คุณภาพต่ำ และบริบทไม่เพียงพอ เพียงความไม่แน่นอนของข้อมูลในระดับปานกลางก็สามารถผลักดันความเสี่ยงให้อยู่ในระดับสูงได้อย่างรวดเร็ว

7.4 ความสัมพันธ์ระหว่างปัจจัยต่าง ๆ

จากกราฟทั้ง 4 เส้น จะเห็นได้ว่าไม่มีเส้นใดเส้นหนึ่งที่ถูกกำหนดโดยตัวแปรเพียงตัวเดียว หากแต่ตำแหน่งของแต่ละเส้นเป็นผลรวมของอิทธิพลจาก C, P และ B พร้อมกัน สิ่งนี้สะท้อนข้อสรุปสำคัญของรายงานฉบับนี้ว่า

"AI Hallucination ไม่ได้เกิดจากปัจจัยเดียว
แต่เป็นผลจากหลายปัจจัยที่ทำงานร่วมกัน"

8 การกำหนดระดับความเสี่ยง

เพื่อให้สามารถตีความค่า H ที่ได้จากแบบจำลองได้ง่ายขึ้น รายงานนี้เสนอเกณฑ์เชิงแนวคิดสำหรับแบ่งระดับความเสี่ยงออกเป็น 3 ระดับ ดังนี้

$$0 \leq H < 0.3 \;\Rightarrow\; \text{ความเสี่ยงต่ำ} \tag{13}$$

$$0.3 \leq H < 0.7 \;\Rightarrow\; \text{ความเสี่ยงปานกลาง} \tag{14}$$

$$0.7 \leq H \leq 1 \;\Rightarrow\; \text{ความเสี่ยงสูง} \tag{15}$$

ช่วงค่า H	**ระดับความเสี่ยง**	**ความหมายเชิงแนวคิด**
$[0, 0.3)$	ต่ำ	แบบจำลองมีแนวโน้มให้คำตอบที่น่าเชื่อถือสูง
$[0.3, 0.7)$	ปานกลาง	ควรมีการตรวจสอบเพิ่มเติมก่อนนำไปใช้
$[0.7, 1]$	สูง	มีแนวโน้มสูงที่จะเกิดเนื้อหาคลาดเคลื่อน

ข้อย้ำสำคัญ: เกณฑ์การแบ่งระดับข้างต้นเป็นเกณฑ์ที่กำหนดขึ้นเพื่อการแสดงผลและตีความของแบบจำลองเชิงแนวคิดในรายงานฉบับนี้เท่านั้น **มิใช่มาตรฐานทางการแพทย์ มาตรฐานอุตสาหกรรม หรือเกณฑ์ที่ได้รับการรับรองอย่างเป็นทางการแต่อย่างใด**

9 การวิเคราะห์เชิงคณิตศาสตร์

9.1 พฤติกรรมของ Logistic Function

พิจารณาสมการ (1) ภายใต้ค่า Z ที่แตกต่างกัน

- **เมื่อ Z มีค่าต่ำมาก** (เช่น $Z \to -\infty$) พจน์ e^{-Z} จะมีค่าสูงมาก ทำให้ $H \to 0$ นั่นคือความเสี่ยงเข้าใกล้ศูนย์

- **เมื่อ Z ใกล้ 0** ค่า H จะเข้าใกล้ 0.5 ซึ่งเป็นจุดกึ่งกลางของฟังก์ชัน และเป็นบริเวณที่ความชันของเส้นโค้งมีค่าสูงที่สุด
- **เมื่อ Z มีค่าสูงมาก** (เช่น $Z \to +\infty$) พจน์ e^{-Z} จะเข้าใกล้ศูนย์ ทำให้ $H \to 1$ นั่นคือ ความเสี่ยงเข้าใกล้ค่าสูงสุด

9.2 อนุพันธ์ของฟังก์ชันโลจิสติก

อนุพันธ์ของ H เทียบกับ Z สามารถหาได้ดังนี้ และมีคุณสมบัติพิเศษที่สามารถเขียนในรูปของ H เอง

$$\frac{dH}{dZ} = H(1 - H) \tag{16}$$

ความหมายของอนุพันธ์นี้ในบริบทของแบบจำลองคือ อัตราการเปลี่ยนแปลงของความเสี่ยง H เมื่อ Z เปลี่ยนแปลงไปเล็กน้อย จะมีค่าสูงที่สุดเมื่อ $H = 0.5$ (ซึ่งให้ค่า $\frac{dH}{dZ} = 0.25$ ซึ่งเป็นค่าสูงสุดของฟังก์ชันนี้) และจะมีค่าเข้าใกล้ศูนย์เมื่อ H เข้าใกล้ 0 หรือ 1 กล่าวอีกนัยหนึ่ง คือ ความเสี่ยงของ Hallucination จะไวต่อการเปลี่ยนแปลงของปัจจัยต่าง ๆ มากที่สุดในช่วงความเสี่ยงปานกลาง และจะไม่ไวต่อการเปลี่ยนแปลงมากนักเมื่อความเสี่ยงอยู่ในระดับต่ำมากหรือสูงมากอยู่แล้ว

9.3 ผลของการเปลี่ยนค่าพารามิเตอร์

การเปลี่ยนแปลงค่าสัมประสิทธิ์ a, b, c, d จะส่งผลต่อความชันของเส้นโค้งในทิศทางที่สัมพันธ์กับตัวแปรนั้น ๆ กล่าวคือ ค่าสัมประสิทธิ์ที่มีขนาดสัมบูรณ์มากขึ้นจะทำให้เส้นโค้งมีความชันมากขึ้น (เปลี่ยนแปลงเร็วขึ้น) เมื่อตัวแปรที่เกี่ยวข้องเปลี่ยนแปลง ในขณะที่ค่าคงที่ k มีผลเพียงการเลื่อนตำแหน่งของเส้นโค้งบนแกน x โดยไม่เปลี่ยนรูปทรงของเส้นโค้ง

10 การเชื่อมโยงกับการทดลอง AI

แบบจำลองเชิงแนวคิดนี้สามารถนำไปใช้เป็น **กรอบความคิด (Conceptual Framework)** สำหรับการเปรียบเทียบผลของการออกแบบ Prompt หรือรูปแบบการตั้งคำถามในเชิงคุณภาพได้ เช่น การเปรียบเทียบว่า Prompt ที่มีโครงสร้างชัดเจนกว่าอาจสัมพันธ์กับแนวโน้มความเสี่ยงที่ต่ำกว่าในเชิงแนวคิด หรือการเปรียบเทียบว่าการให้บริบทเพิ่มเติมแก่แบบจำลองอาจสัมพันธ์กับแนวโน้มความเสี่ยงที่ลดลง

อย่างไรก็ตาม การนำแบบจำลองนี้ไปใช้ในลักษณะดังกล่าวต้องอยู่ภายใต้ข้อจำกัดที่ชัดเจนว่า **ค่าตัวเลขที่ได้จากแบบจำลองมิใช่ค่าความเสี่ยงที่วัดได้จริงจากระบบ AI ใด ๆ** หากแต่เป็นเพียงกรอบแนวคิดเชิงคุณภาพที่ช่วยจัดระเบียบความเข้าใจเกี่ยวกับปัจจัยที่อาจเกี่ยวข้องกับ Hallucination เท่านั้น

11 ข้อจำกัดของแบบจำลอง

แบบจำลองที่นำเสนอในรายงานฉบับนี้มีข้อจำกัดที่สำคัญ ดังนี้

1. ค่าสัมประสิทธิ์ a, b, c, d และค่าคงที่ k เป็นค่าที่กำหนดขึ้นเพื่อการสาธิตและจำลองพฤติกรรมของแบบจำลองเท่านั้น
2. ค่าดังกล่าว **ไม่ใช่ข้อมูลที่ได้จากการฝึกฝน (Training) หรือการประเมินผลจริงของระบบ AI ใด ๆ**
3. แบบจำลองนี้ **ไม่สามารถนำไปใช้ทำนายความเสี่ยง Hallucination ที่แท้จริงของแบบจำลองปัญญาประดิษฐ์ใดโมเดลหนึ่งได้**
4. ในระบบ AI จริง Hallucination มีปัจจัยที่เกี่ยวข้องอีกจำนวนมาก เช่น สถาปัตยกรรมของแบบจำลอง วิธีการฝึกฝน คุณภาพและความครอบคลุมของข้อมูลฝึก กลไกการถอดรหัส (Decoding) และปัจจัยเชิงระบบอื่น ๆ ซึ่งแบบจำลองนี้มิได้ครอบคลุมถึง
5. วัตถุประสงค์หลักของแบบจำลองนี้คือการอธิบาย **ความสัมพันธ์เชิงแนวคิด** ระหว่างปัจจัยต่าง ๆ กับความเสี่ยงของ Hallucination มิใช่การสร้างเครื่องมือวัดผลเชิงปริมาณที่แม่นยำ

12 สรุป

รายงานฉบับนี้ได้นำเสนอแบบจำลองเชิงแนวคิดทางคณิตศาสตร์เพื่ออธิบายความเสี่ยงของ AI Hallucination โดยอาศัยฟังก์ชันโลจิสติกเป็นแกนหลักในการรวมปัจจัยสี่ประการ ได้แก่ ความไม่แน่นอนของข้อมูล ความซับซ้อนของคำถาม คุณภาพของ Prompt และความเพียงพอของบริบท เข้าไว้ในสมการเดียวกัน ผลจากการวิเคราะห์กราฟทั้ง 4 กรณีแสดงให้เห็นอย่างชัดเจนว่าความเสี่ยงของ Hallucination มิได้ถูกกำหนดโดยปัจจัยใดปัจจัยหนึ่งเพียงลำพัง หากแต่เป็นผลรวมของปัจจัยหลายด้านที่ทำงานร่วมกันอย่างซับซ้อน

ในเชิงคณิตศาสตร์ คุณสมบัติของฟังก์ชันโลจิสติกและอนุพันธ์ของฟังก์ชันดังกล่าวช่วย อธิบายได้ว่าความเสี่ยงจะมีความไวต่อการเปลี่ยนแปลงของปัจจัยต่าง ๆ มากที่สุดในช่วงความเสี่ยงระดับปานกลาง ซึ่งเป็นช่วงที่มีนัยสำคัญเชิงแนวคิดต่อการออกแบบมาตรการลดความเสี่ยง เช่น การพัฒนาคุณภาพของ Prompt หรือการเพิ่มบริบทที่เกี่ยวข้อง

ท้ายที่สุดนี้ ขอเน้นย้ำอีกครั้งว่าแบบจำลองที่นำเสนอในรายงานฉบับนี้เป็น **เครื่องมือเชิงแนวคิด (Conceptual Model)** ที่มีวัตถุประสงค์เพื่อสนับสนุนความเข้าใจเชิงมโนทัศน์เกี่ยวกับปัจจัยที่อาจเกี่ยวข้องกับ AI Hallucination เท่านั้น มิใช่มาตรวัดทางวิทยาศาสตร์ที่ผ่านการพิสูจน์หรือรับรองจากการทดลองจริงแต่อย่างใด