

ข้อเสนอโครงการฉบับสมบูรณ์ (Full Proposal)

ยุทธศาสตร์	การพัฒนาเศรษฐกิจไทยด้วยเศรษฐกิจสร้างคุณค่าและเศรษฐกิจสร้างสรรค์ ให้มีความสามารถในการแข่งขัน และพึ่งพาตนเองได้อย่างยั่งยืน พร้อมสู่อนาคต โดยใช้วิทยาศาสตร์ การวิจัยและนวัตกรรม
------------	---

ข้อมูลทั่วไป

ชื่อโครงการ (ไทย) โครงการโมเดลปัญญาประดิษฐ์ภาษาไทยขนาดใหญ่: โอเพ่นไทยจีพีที
(อังกฤษ) Open Large Thai Language Model: OpenThaiGPT

ชื่อหัวหน้าโครงการ (ไทย) ดร. กอบกฤตย์ วิริยะยุทธกร

(อังกฤษ) Kobkrit Viriyayudhakorn,. Ph.D.

ตำแหน่ง นายกสมาคมฯ หน่วยงาน สมาคมผู้ประกอบการ
ปัญหาประดิษฐ์ประเทศไทย

สถานที่ติดต่อ 80/359 หมู่ที่ 3 ซ. คลองหลวง 26 ถ. พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง
ปทุมธานี 12120 ประเทศไทย

โทรศัพท์ 086-322-5858

อีเมล kobkrit@aieat.or.th

ชื่อหัวหน้าสถาบัน/ต้นสังกัด ดร. กอบกฤตย์ วิริยะยุทธกร

ตำแหน่ง นายกสมาคมฯ หน่วยงาน สมาคมผู้ประกอบการ
ปัญญาประดิษฐ์ประเทศไทย

สถานที่ติดต่อ 80/359 หมู่ที่ 3 ซ. คลองหลวง 26 ถ. พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง
ปทุมธานี 12120 ประเทศไทย

บทคัดย่อและคำสำคัญ (รวมกันไม่เกิน 1 หน้ากระดาษ A4)

บทคัดย่อ (ไทย)

ในปัจจุบัน การพัฒนา Large Language Model (LLM) สำหรับภาษาต่าง ๆ ได้กลายเป็นสิ่งจำเป็น และมีความสำคัญเพื่อเสริมสร้างประสิทธิภาพในงานด้านปัญญาประดิษฐ์ โดยเฉพาะภาษาไทยที่ยังมีทรัพยากร และข้อมูลจำกัด การพัฒนา LLM จะทำให้ระบบปัญญาประดิษฐ์เข้าใจและประมวลผลภาษาไทยได้อย่างมีประสิทธิภาพ ทำให้สามารถสื่อสารกับผู้ใช้ภาษาไทยได้ง่ายขึ้น

และเปิดโอกาสในการพัฒนาแอปพลิเคชันด้านปัญญาประดิษฐ์ ที่สามารถเข้าใจและตอบสนองต่อความต้องการของผู้ใช้ภาษาไทย เช่น การสืบค้นข้อมูล แปลภาษา และการสร้างข้อความอัตโนมัติ อีกทั้ง LLM ยังช่วยด้านการศึกษาสามารถช่วยให้นักวิจัยและนักศึกษาภาษาไทยสามารถเข้าถึงและใช้ประโยชน์จาก เทคโนโลยีปัญญาประดิษฐ์ นำไปสู่ความก้าวหน้าในงานวิจัยและการสร้างสรรค์

และสุดท้ายช่วยสร้างโอกาสทางเศรษฐกิจและสังคมด้วย LLM ภาษาไทย นักพัฒนาสามารถสร้างแอปพลิเคชัน ที่มีประโยชน์และสามารถขับเคลื่อนธุรกิจและการพัฒนาสังคม ทำให้เกิดสัมพันธภาพที่ยั่งยืน

การพัฒนาโมเดลภาษาไทยขนาดใหญ่นั้น เริ่มจากการรวบรวมและจัดการข้อมูล (data preprocessing) ที่สอดคล้องกับภาษาไทย หากเปรียบเทียบกับภาษาอื่น ๆ ซึ่งมีความซับซ้อนและ เฉพาะเจาะจง เช่น การแบ่งคำ และปัญหาการใช้วรรณยุกต์ ต่อมาคือการเลือกโครงสร้างของโมเดล (model architecture) ที่เหมาะสมสำหรับภาษาไทย พิจารณาความสามารถในการเรียนรู้ลักษณะเฉพาะ และปรับปรุง ประสิทธิภาพของโมเดล โดยใช้ Transfer Learning หรือ Reinforcement Learning with Human Feedback (RLHF) และการปรับปรุงค่า Hyperparameters นอกจากนี้ยังควรมีการวัดผลและประเมินความ แม่นยำของโมเดล (model evaluation) ด้วยเครื่องมือที่เหมาะสม เช่น BLEU, ROUGE, และ F1 Score เพื่อให้เห็นถึงความสามารถในงานต่าง ๆ ที่เกี่ยวข้องกับภาษาไทย

ในที่สุด การพัฒนา LLM สำหรับภาษาไทยในรูปแบบ Opensource ไม่เพียงแต่ส่งเสริมความร่วมมือ และการแลกเปลี่ยนความรู้ในชุมชนวิชาการและนักพัฒนา แต่ยังเปิดโอกาสให้ภาษาไทยได้รับประโยชน์จาก ปัญญาประดิษฐ์และนวัตกรรมใหม่ ๆ ในอนาคต

บทคัดย่อ (อังกฤษ)

In today's world, developing Large Language Models (LLMs) for various languages has become essential and important to enhance the effectiveness of artificial intelligence tasks, particularly for Thai language, which still has limited resources and data. Developing LLMs for the Thai language would enable AI systems to understand and process the language more efficiently, making communication with Thai users easier, and opening opportunities for the development of AI applications that can understand and respond to the needs of Thai users, such as information retrieval, language translation, and automatic text generation. Additionally, LLMs can support education and help Thai researchers and students access and benefit from AI technologies, leading to progress in research and creative work. Lastly, it creates economic and social opportunities with Thai LLMs, allowing developers to create

useful applications that can drive business and social development, resulting in sustainable partnerships.

Developing a large Thai language model starts with collecting and preprocessing data that is compatible with the Thai language, which is more complex and specific compared to other languages, such as word segmentation and issues related to diacritics usage. Next is choosing an appropriate model architecture for the Thai language, considering the ability to learn specific features, and improving the model's performance using Transfer Learning or Reinforcement Learning with Human Feedback (RLHF) and adjusting hyperparameters. Moreover, it is crucial to evaluate the model's accuracy and performance using suitable tools such as BLEU, ROUGE, and F1 Score to demonstrate the capabilities in various tasks related to the Thai language.

Ultimately, developing an open-source LLM for the Thai language not only promotes collaboration and knowledge exchange in the academic and developer communities but also opens up opportunities for the Thai language to benefit from AI and future innovations.

คำสำคัญ (ไทย): โมเดลปัญญาประดิษฐ์ภาษาขนาดใหญ่, ภาษาไทย, การเรียนรู้ด้วยเครื่อง, เทคนิคทรานส์ฟอร์มเมอร์, การเรียนรู้ด้วยตัวอย่างเพียงเล็กน้อย, โครงการพื้นฐานด้านปัญญาประดิษฐ์ของประเทศ, ปัญญาประดิษฐ์ทั่วไป

คำสำคัญ (อังกฤษ): Large Language Model, Thai, Machine Learning, Transformer, Few-shot Learning, Nation AI Infrastructure, AGI

รายละเอียดของคณะผู้วิจัย

	รายชื่อคณะผู้วิจัย	% เวลาที่ใช้ในการทำโครงการ*	% ความรับผิดชอบในโครงการ**
(1)	ชื่อหัวหน้าโครงการ ดร.กอบกฤตย์ วิริยะยุทธกร	40.00%	33.33%
	สังกัด สมาคมผู้ประกอบการปัญญาประดิษฐ์ประเทศไทย		
	ความเชี่ยวชาญ การพัฒนาธุรกิจสินค้าและบริการ ปัญญาประดิษฐ์ / การบริหารวิสาหกิจใหม่ (Startup) / Natural Language Processing		
(2)	ชื่อผู้ร่วมโครงการ ดร. สุเมธ ยืนยง	40.00%	33.33%
	สังกัด คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล		
	ความเชี่ยวชาญ Deep Learning / Machine Learning / Transformer / Language Model		
(3)	ชื่อผู้ร่วมโครงการ ดร. ปรัชญา บุญขวัญ	40.00%	33.33%
	สังกัด ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (เนคเทค)		
	ความเชี่ยวชาญ Deep Learning / Machine Learning / Transformer / Language Model		
			100 %

ผลิตภัณฑ์ เทคโนโลยีเทียบเคียง / คู่แข่ง (ระบุข้อมูลเปรียบเทียบผลิตภัณฑ์ เทคโนโลยีเทียบเคียง / คู่แข่งในปัจจุบันทั้งทางตรงและทางอ้อม)

หัวข้อเปรียบเทียบ (เช่น ด้านประสิทธิภาพ คุณภาพ ฟังก์ชันการใช้งาน กระบวนการผลิต ต้นทุน ความปลอดภัย ฯลฯ โดย สามารถปรับ-เพิ่มหัวข้อได้ตามความ เหมาะสม)	ผลิตภัณฑ์ / เทคโนโลยี ของ ท่าน OpenThaiGPT	คู่แข่งลำดับ 1 (ระบุชื่อ) OpenAI's ChatGPT	คู่แข่งลำดับ 2 (ระบุชื่อ) Facebook's Llama	คู่แข่งลำดับ 3 (ระบุชื่อ) VISTEC's Wangchanberta
โมเดลภาษาขนาดใหญ่	1.3 – 10 Billion Parameters	175 Billion parameters	7 Billion parameters	340 million parameters
สนับสนุนภาษาไทย	ใช่	ไม่	ไม่	ไม่
ทำตามคำสั่ง (Instructable)	ใช่	ใช่	ใช่	ไม่
เขียนต่อได้ (Generative)	ใช่	ใช่	ใช่	ไม่
เรียนรู้จากตัวอย่างเล็กน้อย (Few Shot Learning)	ใช่	ใช่	ใช่	ไม่

ข้อมูลด้านตลาด / ผู้ใช้

1.1 ขนาดและแนวโน้มตลาด

1.1.1 ภาพรวมของตลาด

ขนาดตลาด (Market size):

ตลาด NLP ทั่วโลก: \$26.42 billion in 2022 (870,000 ล้านบาท)

ตลาด NLP ในไทยใน ปี 2022 (เทียบกับสัดส่วน GDP ระหว่างไทยและทั่วโลก 0.6%):

870,000 ล้านบาท \times 0.6% = 5,220 ล้านบาท

Year	Market Size (Million THB)
2022	5,220.00
2023	6,164.82
2024	7,280.65
2025	8,598.45
2026	10,154.77
2027	11,992.78
2028	14,163.48
2029	16,727.07
2030	19,754.67

อัตราการเติบโต (CAGR): 18.1% => จะมีมูลค่า 2
หมื่นล้านบาท ภายในปี 2030

การวิเคราะห์คู่แข่ง (Competitor analysis):

- OpenAI มุ่งเน้นทำ AI as a Service (AaaS) ผ่าน API สำหรับ 20 ภาษาหลักทั่วโลก โดยไม่ได้มีภาษาไทยเป็นเป้าหมายหลัก ไม่เข้าถึงตลาด ที่ต้องใช้ Opensource Model (อาทิ เช่น ตลาด Offline Automation, Embedded System) ที่ใช้ในการทำงานเป็นหลัก
- OpenAI เก็บค่าใช้จ่ายสูงถึง 7 สตางค์ต่อ 1,000 โทเคน หรือ 500 ตัวอักษรไทย (2 โทเคน ต่อ 1 ตัวอักษรไทย) ทำให้ Service ที่พึ่งพาภาษาไทยเป็นหลัก มีต้นทุนในการใช้ LLM ของ OpenAI มีต้นทุนที่สูง
- Microsoft มุ่ง Enterprise LLM ซึ่งมีราคาสูงกว่า OpenAI เสียอีก และใช้ Technology เดียวกับ OpenAI
- Google ยังเน้นพัฒนาภาษาอังกฤษเป็นหลักอย่างเดียว เพื่อสู้กับ OpenAI

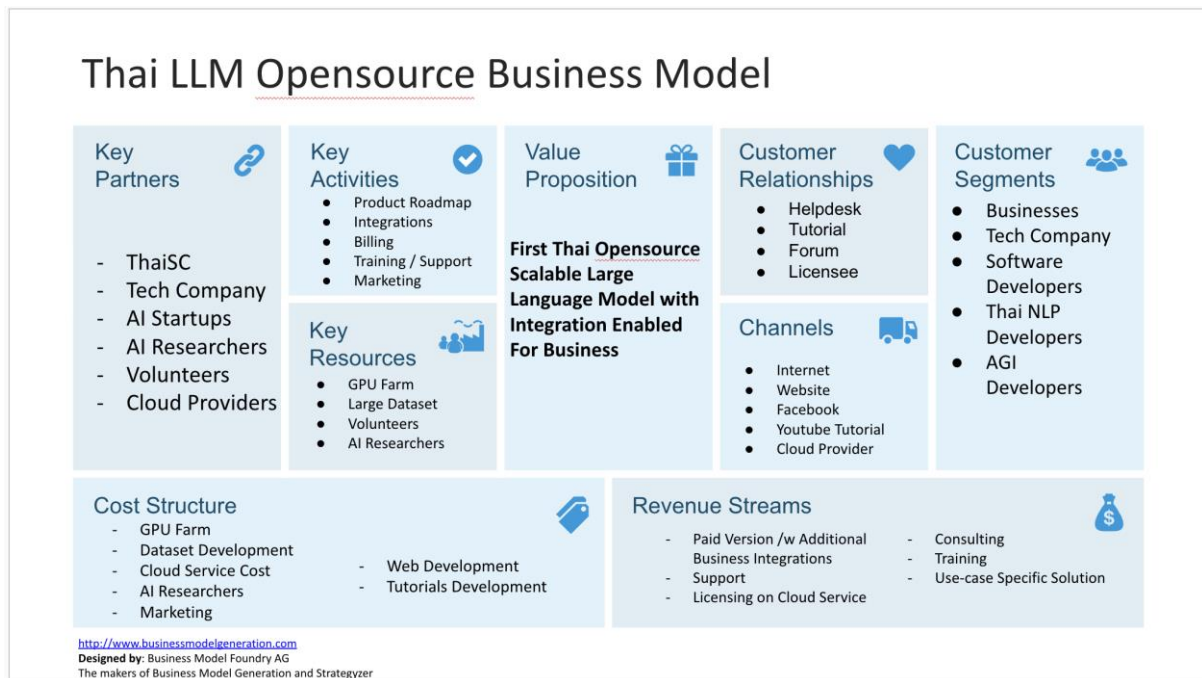
ตลาดที่คาดว่าจะสามารถเข้าถึง (Market share):

ต้นทุนด้าน LLM Service ที่นักพัฒนายอมจ่าย เพื่อนำไปทำ Product และ Service มักไม่ควรเกิน 10% ของมูลค่าทั้งหมด หรือประมาณ 500 ล้านบาทในปี 2023 และจะเป็น 2,000 ล้านบาทในปี 2030 Market Share ที่คาดว่าจะได้รับการโครงการนี้ จะเพิ่มขึ้นเรื่อยๆดังต่อไปนี้

Year	Market Size (Million THB)
2022	522.00
2023	616.48
2024	728.07
2025	859.85
2026	1,015.48
2027	1,199.28
2028	1,416.35
2029	1,672.71
2030	1,975.47

<https://www.fortunebusinessinsights.com/industry-reports/natural-language-processing-nlp-market-101933>

แผนธุรกิจ (ระบุข้อมูลรายละเอียดซึ่งประกอบด้วยกลุ่มลูกค้าเป้าหมาย ราคาของผลิตภัณฑ์ที่พัฒนา ช่องทางการจัดจำหน่าย ฯลฯ)



ข้อมูลความต้องการของผู้ใช้ / ลูกค้า (ถ้ามี) (ระบุข้อมูลรายละเอียดซึ่งเป็นผลจากหาหรือหรือสัมภาษณ์ผู้ใช้หรือผลการทดสอบใช้งานผลิตภัณฑ์หรือบริการร่วมกับผู้ใช้ (Users) ในช่วงที่ผ่านมา)

Pains

สัมภาษณ์จากผู้ที่ใช้ ChatGPT ภาษาไทยจำนวน 10 คน

1. ใช้ OpenAI's ChatGPT ภาษาไทยตอบช้า ตอบได้สั้น - 10 ท่าน
2. ราคาสูง (7 สตางค์ต่อ 500 ตัวอักษร) - 8 ท่าน
3. ไม่มีเสถียรภาพ ล่มบ่อย - 5 ท่าน
4. ไม่สามารถเชื่อมต่อ ChatGPT กับองค์ความรู้ของ Business ของตัวเองได้ - 3 ท่าน
5. ไม่สามารถปรับปรุงแก้ไข ChatGPT ได้ด้วยตัวเอง - 1 ท่าน

หน่วยงาน /บริษัทร่วมทุน

ชื่อนิติบุคคล ภาษาไทย...บริษัท ไอแอปพ์เทคโนโลยี จำกัด

ชื่อนิติบุคคล ภาษาอังกฤษ...iApp Technology Co., Ltd.

วันที่จดทะเบียน 8/8/2013 ทะเบียนเลขที่ 0135556015413

ก่อตั้งเมื่อ: 2013...ทุนจดทะเบียน: 21.5 ล้านบาท

ทุนจดทะเบียนที่เรียกชำระแล้ว: 21.5 ล้าน บาท เมื่อวันที่: 1 กุมภาพันธ์ 2565

สัดส่วนผู้ถือหุ้นสัญชาติไทย (%).....100%..... สัญชาติอื่น (%)

ที่อยู่กิจการ..80/359 หมู่ที่ 3 ซ. คลองหลวง 26 ถ. พหลโยธิน ต.คลองหนึ่ง อ.คลองหลวง ปทุมธานี

12120 โทรศัพท์.....02-051-2453

1.2 ประวัติความเป็นมาของบริษัทโดยย่อ

บริษัท ไอแอฟเทคโนโลยี จำกัด ก่อตั้งเมื่อ วันที่ 8 สิงหาคม พ.ศ. 2556 โดยผู้ก่อตั้ง คือ ดร. กอบกฤตย์ วิริยะยุทธกร โดยสำนักงานใหญ่ตั้งอยู่เลขที่ 80/359 หมู่ที่ 3 ซอยคลองหลวง 26 ตำบลคลองหนึ่ง อำเภอคลองหลวง จังหวัดปทุมธานี

ในช่วงปีแรก (พ.ศ. 2556 - พ.ศ. 2558) บริษัท ไอแอฟเทคโนโลยี จำกัด ตั้งอยู่เลขที่ 699 ถนนเจริญนคร แขวงคลองตันใต้ เขตคลองสาน จังหวัดกรุงเทพมหานคร มีเป้าหมายในการก่อตั้งเป็นบริษัทที่รับพัฒนา Website และ Mobile Application โดยมุ่งเน้นการพัฒนาด้วยเทคโนโลยีภาษา JavaScript เป็นหลัก เนื่องจากการเป็นครั้งเดียวแล้วสามารถนำไปใช้ได้ทั้ง Website และ Mobile Application ซึ่งเป็นการลดขั้นตอนการพัฒนาโปรแกรม และได้ผลงานที่มีคุณภาพสูงตอบโจทย์การใช้งานของลูกค้าได้เป็นอย่างดี รวมถึงการบำรุงรักษาโปรแกรกดังกล่าว (Maintenance) ดำเนินการได้ง่ายในระยะเวลาอันสั้น และรวดเร็ว อีกทั้งยังสามารถช่วยลดต้นทุนค่าดำเนินการได้เป็นอย่างดี

ต่อมาในปี พ.ศ. 2559 บริษัท ฯ ได้ย้ายสำนักงานใหญ่ มาอยู่ที่ตั้งของสำนักงานปัจจุบัน เพื่อรองรับจำนวนที่เพิ่มขึ้นของบุคลากร ที่ผ่านการคัดเลือกจากคุณสมบัติ ความสามารถ และประสบการณ์การทำงาน เพื่อเตรียมความพร้อมสำหรับการพัฒนา Website และ Mobile Application ให้กับโครงการทั้งในส่วนราชการ รัฐวิสาหกิจ องค์กรมหาชน สถาบันการศึกษา และภาคเอกชน ที่มีความแตกต่าง ชับซ้อน ให้มีประสิทธิภาพและเป็นมาตรฐานเดียวกัน

ต่อมาในปี พ.ศ. 2561 บริษัท ฯ เริ่มสนใจงานด้านการพัฒนาหุ่นยนต์สนทนาอัตโนมัติ (Chatbot) และปัญญาประดิษฐ์ (Artificial Intelligence: AI) ภาษาไทยมากยิ่งขึ้น โดยเล็งเห็นว่างานดังกล่าวเป็นแนวทางของนวัตกรรมที่สำคัญในอนาคต ประกอบกับช่วงเวลานั้นยังไม่มีผู้ใดทุ่มเท และให้ความสนใจเกี่ยวกับการพัฒนาสำหรับภาษาไทย เนื่องจากภาษาไทยมีโครงสร้างซับซ้อน และมีความเป็นท้องถิ่นสูง แต่ในทางกลับกันบริษัท ฯ มีความมุ่งมั่น และเชื่อมั่นว่าจะสามารถพัฒนานวัตกรรมดังกล่าวได้เป็นอย่างดี เทียบเท่าบริษัทต่างประเทศได้ เนื่องจากผู้ก่อตั้งบริษัท ฯ มีความรู้ ความชำนาญ และความเชี่ยวชาญเฉพาะทางด้านการพัฒนาหุ่นยนต์สนทนาอัตโนมัติ (Chatbot) และปัญญาประดิษฐ์ (Artificial Intelligence: AI) ภาษาไทยเป็นอย่างดี ต่อมา ในปีเดียวกันช่วงเดือนเดือนพฤศจิกายน (พ.ศ. 2561) บริษัท ฯ ได้รับเงินทุนสนับสนุนจาก บริษัท จันฉนวนชัย ซีเคียวริตี้ พรินท์ติ้ง จำกัด เป็นจำนวน 12 ล้านบาท เพื่อเป็นจุดเริ่มต้นการลงทุนที่ต้องการมุ่งเน้นการพัฒนาปัญญาประดิษฐ์(Artificial Intelligence: AI) ภาษาไทย อย่างจริงจัง

ต่อมาในปี พ.ศ. 2562 บริษัท ฯ ได้เริ่มเปิดตัวผลิตภัณฑ์และบริการ ที่พัฒนาต่อยอดจากระบบปัญญาประดิษฐ์ (Artificial Intelligence: AI) ภาษาไทยดังต่อไปนี้

- 1) หุ่นยนต์บริการ (Robot Service)
- 2) ระบบสนทนาโต้ตอบอัตโนมัติภาษาไทย (Chatbot)
- 3) ปัญญาประดิษฐ์สำหรับอ่านข้อมูลหน้าบัตรประชาชน (National ID Card OCR)
- 4) บริการด้านปัญญาประดิษฐ์ด้านอื่นๆ (AI Services)

บริษัท ฯ มีแนวความคิด และนโยบายในการให้ความสำคัญด้านการพัฒนาผลิตภัณฑ์และบริการที่มีมาตรฐานสูงสุด เพื่อให้ถูกนำไปใช้ในองค์กร หรือหน่วยงานภายในประเทศ ให้ได้รับประโยชน์ และเกิดประสิทธิภาพ โดยจากความมุ่งมั่นที่ยึดถือนโยบายอย่างเคร่งครัด ได้ส่งผลให้บริษัท ไอแอฟพี เทคโนโลยี จำกัด คือ หนึ่งในผู้นำทางด้านหุ่นยนต์บริการภาษาไทย (Service Robot) หุ่นยนต์สนทนาอัตโนมัติ (Chatbot) ปัญญาประดิษฐ์ (Artificial Intelligence: AI) ภาษาไทย ระบบปัญญาประดิษฐ์สำหรับอ่านข้อมูลหน้าบัตรประชาชน (National ID Card OCR) และระบบ AI Service Marketplace ของคนไทย

1.3 จุดเด่น ข้อได้เปรียบของผู้ร่วมทุน (ระบุจุดเด่น ข้อได้เปรียบที่แสดงให้เห็นว่าผู้ร่วมทุนมีศักยภาพ และขีดความสามารถในการดำเนินงานจนสำเร็จ และนำผลงานไปใช้ได้จริงเป็นข้อ ๆ)

- มีบุคลากรด้านปัญญาประดิษฐ์มากกว่า 15 ท่าน
- มีตลาด AI Service พร้อมให้บริการ (มีการขายบริการ AI API ได้สูงถึง 900,000 ครั้งต่อเดือน)
- มีเงินทุนในการลงทุนและพัฒนา

1.4 ผู้ใช้ประโยชน์จริงเชิงพาณิชย์และแนวทางการใช้ประโยชน์ (พิจารณาทั้ง supply chain อาจมีมากกว่า 1 ราย)

ผู้ใช้ประโยชน์จริง	แผน/แนวทางการใช้ประโยชน์
<input checked="" type="checkbox"/> เอกชนผู้ร่วมทุน	<ul style="list-style-type: none"> ● พัฒนาต่อยอด OpenThaiGPT ให้เป็นระบบ Chatbot อัจฉริยะ ● ให้บริการ OpenThaiGPT As a Service
<input checked="" type="checkbox"/> ผู้ผลิตรายอื่นใน supply chain	<ul style="list-style-type: none"> ● ปลดปล่อยเป็น Opensource ทั้งหมด ให้นำไปใช้งานได้ ● สามารถนำไปต่อยอดเป็นธุรกิจอื่นๆได้
<input checked="" type="checkbox"/> ผู้ใช้รายอื่น	<ul style="list-style-type: none"> ● ปลดปล่อยเป็น Opensource ทั้งหมด ให้นำไปใช้งานได้ ● สามารถนำไปต่อยอดเป็นธุรกิจอื่นๆได้

5.7 ความเป็นไปได้ในการใช้ประโยชน์เชิงพาณิชย์ (commercialization)

ภาคเอกชน (ผู้ให้ทุนร่วม/ผู้รับทุน) มีงบประมาณพร้อมต่อการลงทุนในการผลิตเชิงพาณิชย์	<input checked="" type="checkbox"/> ใช่	<input type="checkbox"/> ไม่ใช่	<input type="checkbox"/> กำลังหาข้อมูล
ภาคเอกชน (ผู้ให้ทุนร่วม/ผู้รับทุน) มีแผนธุรกิจสำหรับการผลิตเชิงพาณิชย์	<input checked="" type="checkbox"/> ใช่	<input type="checkbox"/> ไม่ใช่	<input type="checkbox"/> กำลังหาข้อมูล
ภาคเอกชนรายอื่นมารับถ่ายทอดเทคโนโลยี	<input checked="" type="checkbox"/> ใช่	<input type="checkbox"/> ไม่ใช่	<input type="checkbox"/> กำลังหาข้อมูล
การออกไปจัดตั้งบริษัท (spin off)	<input checked="" type="checkbox"/> ใช่	<input type="checkbox"/> ไม่ใช่	<input type="checkbox"/> กำลังหาข้อมูล
มีแผนหรือ roadmap การขอขึ้นทะเบียน อย. หรือรับรองมาตรฐาน	<input checked="" type="checkbox"/> ใช่	<input type="checkbox"/> ไม่ใช่	<input type="checkbox"/> กำลังหาข้อมูล
ผลิตภัณฑ์สามารถออกสู่ตลาดได้ภายใน	<input type="checkbox"/> เร็วกว่า 2 ปี <input type="checkbox"/> 2-5 ปี <input checked="" type="checkbox"/> มากกว่า 5 ปี <input type="checkbox"/> กำลังหาข้อมูล		

ปริมาณวัตถุดิบเพียงพอสำหรับการผลิตเชิงพาณิชย์	<input checked="" type="checkbox"/> ใช่	<input type="checkbox"/> ไม่ใช่	<input type="checkbox"/> กำลังหาข้อมูล
แหล่งวัตถุดิบเพื่อการผลิตเชิงพาณิชย์มาจาก	<input checked="" type="checkbox"/> ผลิตในประเทศ <input type="checkbox"/> นำเข้าจากต่างประเทศ <input type="checkbox"/> กำลังหาข้อมูล		

หลักการ ที่มา และความสำคัญของปัญหาวิจัย

ในปัจจุบัน การพัฒนา Large Language Model (LLM) สำหรับภาษาต่าง ๆ ได้กลายเป็นสิ่งจำเป็นและมีความสำคัญเพื่อเสริมสร้างประสิทธิภาพในงานด้านปัญญาประดิษฐ์ โดยเฉพาะภาษาไทยที่ยังมีทรัพยากรและข้อมูลจำกัด การพัฒนา LLM จะทำให้ระบบปัญญาประดิษฐ์เข้าใจและประมวลผลภาษาไทยได้อย่างมีประสิทธิภาพ ทำให้สามารถสื่อสารกับผู้ใช้ภาษาไทยได้ง่ายขึ้นและเปิดโอกาสในการพัฒนาแอปพลิเคชันด้านปัญญาประดิษฐ์ ที่สามารถเข้าใจและตอบสนองต่อความต้องการของผู้ใช้ภาษาไทย เช่น การสืบค้นข้อมูล แปลภาษา และการสร้างข้อความอัตโนมัติ อีกทั้ง LLM ยังช่วยด้านการศึกษามารถช่วยให้นักวิจัยและนักศึกษาภาษาไทยสามารถเข้าถึงและใช้ประโยชน์จากเทคโนโลยีปัญญาประดิษฐ์ นำไปสู่ความก้าวหน้าในงานวิจัยและการสร้างสรรค์และสุดท้ายช่วยสร้างโอกาสทางเศรษฐกิจและสังคมด้วย LLM ภาษาไทย นักพัฒนาสามารถสร้างแอปพลิเคชันที่มีประโยชน์และสามารถขับเคลื่อนธุรกิจและการพัฒนาสังคม ทำให้เกิดสัมพันธภาพที่ยั่งยืน

LLM ภาษาไทยที่เปิดแบบ Opensource ซึ่งเป็น Building Block สำคัญของการพัฒนาด้านปัญญาประดิษฐ์ของประเทศไทยยังไม่มีการพัฒนา การพัฒนา LLM ภาษาไทยที่เปิดกว้างให้เข้าถึงได้ง่ายจะช่วยให้ความมั่นคงด้านปัญญาประดิษฐ์ของไทย เป็นผู้นำของอาเซียน และได้เก็บเกี่ยวผลประโยชน์จากการเป็นผู้นำด้านปัญญาประดิษฐ์ด้านภาษาบนยุคการปฏิวัติปัญญาประดิษฐ์ที่จะเป็นตัวสนับสนุน GDP หลักของประเทศในอีก 10 ปีข้างหน้า

เป้าหมายของงานวิจัย (NRIIS เขียนว่า กรอบการวิจัย)

- ทำการวิจัยในด้านต่างๆดังนี้
 - การพัฒนา Large Language Models (LLMs) และ GPT ภาษาไทย
 - แก้ไขปัญหาความซับซ้อนของภาษาไทย
 - การสร้างและปรับปรุงแบบจำลองภาษาไทย
 - การประเมินประสิทธิภาพของ LLM ภาษาไทย
 - การใช้งานและประโยชน์ของ LLM ภาษาไทย
 - การพัฒนาและแชร์ LLM ภาษาไทยแบบเปิด (Opensource)

วัตถุประสงค์

พัฒนา OpenThaiGPT ในรูปแบบ Large Language Model ออกมาให้สามารถ

- ตอบในบทบาทผู้ช่วย ในรูปแบบภาษาไทยและภาษาอังกฤษได้
- สามารถเรียนรู้การทำสิ่งใหม่ โดยการแสดงตัวอย่างเบื้องต้น (Few-shot Learning) ได้
- สามารถเชื่อมต่อกับองค์ความรู้ภายนอก หรือให้เจ้าของธุรกิจ สอนข้อมูลเพิ่มเติมได้ (External Integration)
- สามารถมีองค์ความรู้ด้านประเทศไทย และใช้ภาษาไทยจนสามารถทำข้อสอบ O-Net ได้เบื้องต้น

แนวคิด ทฤษฎี และสมมติฐานงานวิจัย

พื้นฐานของ Large Language Models (LLMs) และ GPT การพัฒนาและประยุกต์ใช้ Large Language Models (LLMs) ได้กลายเป็นทิศทางสำคัญในวงการปัญญาประดิษฐ์ (AI) โดยเฉพาะในการประมวลผลภาษาธรรมชาติ (NLP) (Brown et al., 2020; Devlin et al., 2019) ตัวอย่างหนึ่งของ LLM คือ GPT (Generative Pre-trained Transformer) ซึ่งเป็นรุ่นของ Transformer ที่ถูกพัฒนาโดย OpenAI (Radford et al., 2018)

ภาษาไทยมีความซับซ้อนในด้านการแบ่งคำและการใช้วรรณยุกต์ เรียกว่าการตัดคำภาษาไทย (Jasabutr & Sornlertlamvanich, 2014) นอกจากนี้ยังมีปัญหาเชิงโครงสร้างภาษา เช่น การประสานคำ (Manurung et al., 2012) ซึ่งต้องคำนึงถึงเพื่อประสิทธิภาพของ LLM ในภาษาไทย การสร้าง LLM ภาษาไทยต้องเริ่มจากการเก็บข้อมูลและประมวลผลเพื่อให้เหมาะสมกับภาษาไทย (Jitkrittum et al., 2017) จากนั้นเลือกโครงสร้างของแบบจำลองที่เหมาะสม เช่น GPT และปรับปรุงประสิทธิภาพของแบบจำลองด้วยการใช้ Transfer Learning (Pan & Yang, 2010) หรือ Reinforcement Learning with Human Feedback (RLHF) (Christiano et al., 2017) และปรับค่าพารามิเตอร์ต่างๆ สำหรับให้เหมาะสมกับภาษาไทย เช่น การปรับค่า learning rate หรือจำนวนชั้นของ Transformer (Vaswani et al., 2017)

การประเมินประสิทธิภาพของ LLM ภาษาไทย เมื่อพัฒนาแบบจำลองภาษาไทยแล้ว จำเป็นต้องประเมินประสิทธิภาพของแบบจำลอง โดยใช้เครื่องมือที่เหมาะสม เช่น BLEU (Papineni et al., 2002) สำหรับการประเมินการแปลภาษา, ROUGE (Lin, 2004) สำหรับการสรุปข้อความ, และ F1 Score (Rijsbergen, 1979) สำหรับการวัดประสิทธิภาพในงานที่ต้องการ

การใช้งานและประโยชน์ของ LLM ภาษาไทย การพัฒนาและปรับปรุง LLM ภาษาไทยจะสามารถนำไปประยุกต์ใช้ในหลายๆ งาน เช่น การค้นหาข้อมูล, การแปลภาษา, การสร้างข้อความอัตโนมัติ, การสนับสนุนการศึกษา, และการช่วยเสริมความก้าวหน้าในงานวิจัยและงานสร้างสรรค์ นอกจากนี้ยังสามารถสร้างโอกาสทางเศรษฐกิจและสังคมโดยการให้นักพัฒนาสร้างแอปพลิเคชันที่มีประโยชน์และขับเคลื่อนการพัฒนาทางธุรกิจและสังคม สร้างและเผยแพร่ LLM ภาษาไทยในรูปแบบเปิด (open-source) ไม่เพียงแต่จะส่งเสริมความร่วมมือและการแลกเปลี่ยนความรู้ในชุมชนวิชาการและนักพัฒนา แต่ยังเปิดโอกาสให้ภาษาไทยได้รับประโยชน์จากการนำ AI และนวัตกรรมใหม่ ๆ ไปประยุกต์ใช้ในหลาย ๆ ด้าน เช่น การประยุกต์ใช้ LLM ภาษาไทยในการสร้างแอปพลิเคชันสำหรับสนับสนุนการเรียนรู้ หรือการวิจัยที่มีความท้าทายทางด้านภาษาไทย

References

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Agarwal, S. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
- Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., & Song, D. (2021). The Secret Sharer: Evaluating and testing unintended memorization in neural networks. *Journal of Privacy and Confidentiality*, 11(1).
- Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems* (pp. 4299-4307).
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Jasabutr, T., & Sornlertlamvanich, V. (2014). Thai word segmentation using maximum matching based on longest common prefix. In *2014 International Conference on Asian Language Processing (IALP)* (pp. 164-167). IEEE.
- Jitkrittum, W., Boonpok, C., & Watcharawittayakul, S. (2017). A Linear-Time Kernel Goodness-of-Fit Test. arXiv preprint arXiv:1705.07673.

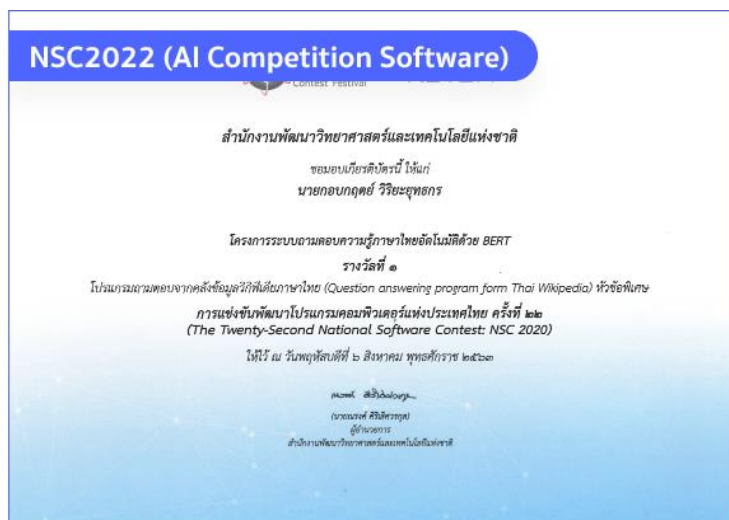
- Lin, C. Y. (2004). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).
- Manurung, R., Ritchie, G., & Thompson, H. (2012). Using genetic algorithms to create meaningful poetic text. Journal of Experimental & Theoretical Artificial Intelligence, 24(1), 43-64.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2021). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220-229).
- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318). Association for Computational Linguistics.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.
- Rijsbergen, C. J. V. (1979). Information retrieval. Butterworth-Heinemann.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

ผลงานเดิม และความเป็นเจ้าของในทรัพย์สินทางปัญญาที่เกี่ยวข้องกับงานวิจัยในโครงการที่จะขอรับทุนนี้

- OpenThaiGPT 0.0.4 โมเดลภาษาภาษาไทยขนาดใหญ่จำนวน 120 Million Parameters ที่เปิด Opensource (Apache 2.0)
<https://openthaigpt.aieat.or.th/released-openthaigpt-pip-python-library-less-than-0.0.10-greater-than>
- ChochaeGPT ระบบแชทบอทปัญญาประดิษฐ์ภาษาไทย ที่สามารถปรับปรุงองค์ความรู้ได้ ต่อยอดบนเทคโนโลยี OpenAI's GPT-4
<https://chochaegpt.iapp.co.th>

- iApp's QA ระบบสกัดคำถามคำตอบภาษาไทย (QA) อัตโนมัติ

https://ai.iapp.co.th/product/thai_automatic_qa ได้รับรางวัลชนะเลิศ NSC2020



- iApp's QG ระบบสร้างคำถามอัตโนมัติ จากบทความ (QG) อัตโนมัติ

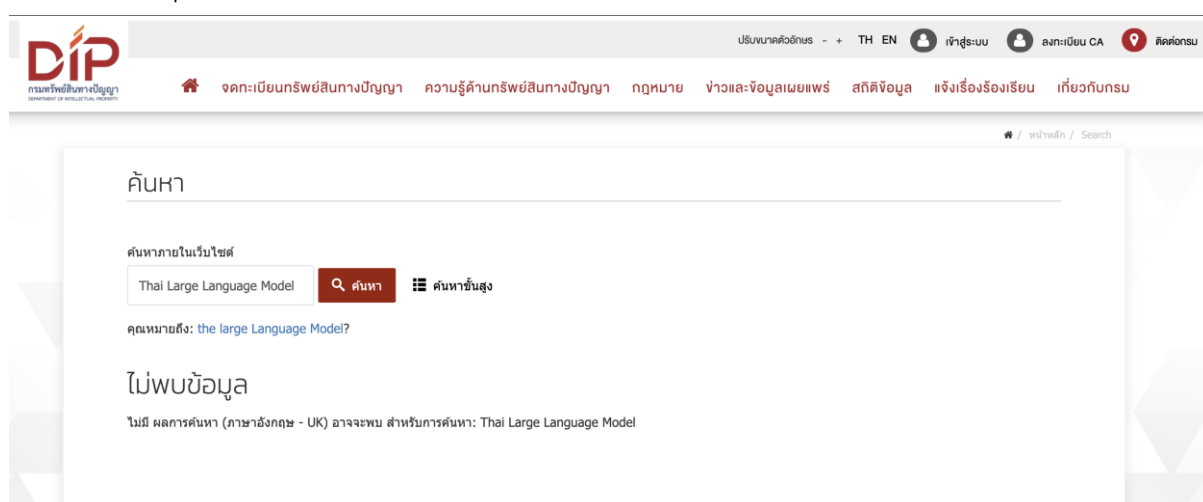
https://ai.iapp.co.th/product/thai_qa_generation

- iApp's Thai Text Summarization : AI ย่อความภาษาไทยแบบเขียนขึ้นมาใหม่ได้เอง (Generative) อัตโนมัติ

https://ai.iapp.co.th/product/thai_text_summarization

การตรวจสอบทรัพย์สินทางปัญญาที่เกี่ยวข้องกับงานวิจัย

ไม่พบข้อมูลทรัพย์สินทางปัญญาที่เกี่ยวข้องกับงานวิจัยในด้าน Large Language Model ภาษาไทยใดๆ



มาตรฐานและกฎหมายที่เกี่ยวข้องกับงานวิจัย และแผนการนำผลิตภัณฑ์หรือบริการที่ได้จากงานวิจัยไปสู่มาตรฐานดังกล่าว

อยู่ในกลุ่มมาตรฐานด้านปัญญาประดิษฐ์ ISO/IEC J42 ดังนี้

ISO/IEC TS 4213:2022

Information technology — Artificial intelligence — Assessment of machine learning classification performance

ISO/IEC 22989:2022

Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

ISO/IEC 23053:2022

Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)

ISO/IEC 23894:2023

Information technology — Artificial intelligence — Guidance on risk management

ISO/IEC TR 24027:2021

Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making

ISO/IEC TR 24028:2020

Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence

ISO/IEC TR 24029-1:2021

Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview

ISO/IEC TR 24368:2022

Information technology — Artificial intelligence — Overview of ethical and societal concerns

ISO/IEC TR 24372:2021

Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems

ISO/IEC 38507:2022

Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations

ระเบียบวิธีดำเนินการวิจัย และแผนการดำเนินงานวิจัย

คำถาม: การสร้าง LLM ภาษาไทยที่มีความสามารถตอบโต้กับมนุษย์ได้อย่างมีประสิทธิภาพ
ทำตามคำสั่ง ตอบคำถาม และแก้ไขปัญหาได้ (Instruction followed) ได้อย่างไร







วิธีการดำเนินแผนการวิจัย

การสร้าง LLM ภาษาไทยที่สามารถตอบโต้กับมนุษย์ได้อย่างมีประสิทธิภาพ

- ต้องเริ่มจากการเก็บข้อมูลและประมวลผลเพื่อให้เหมาะสมกับภาษาไทย (Jitkittum et al., 2017) (Data set Preparing)
- พัฒนาโครงสร้างของแบบจำลองด้านภาษาที่เหมาะสม (Pre-training Model) เช่น GPT และพัฒนา/เทรน Pre-training Model LLM ภาษาไทย
- เตรียมข้อมูลในการตอบคำถาม (Instruction Dataset) สำหรับภาษาไทย
- ปรับปรุงประสิทธิภาพของแบบจำลองด้วยการใช้ Transfer Learning (Pan & Yang, 2010) จาก Instruction Dataset หรือ Reinforcement Learning with Human Feedback (RLHF) (Christiano et al., 2017) (Finetuning)
- การประเมินประสิทธิภาพของ LLM ภาษาไทย เมื่อพัฒนาแบบจำลองภาษาไทย แล้ว จำเป็นต้องประเมินประสิทธิภาพของแบบจำลอง โดยใช้เครื่องมือที่เหมาะสม เช่น BLEU (Papineni et al., 2002) สำหรับการประเมินการแปลภาษา, ROUGE (Lin, 2004) สำหรับการสรุปข้อความ, และ F1 Score (Rijsbergen, 1979) สำหรับการวัดประสิทธิภาพในงานที่ต้องการ (Evaluating)

- ปรับค่าพารามิเตอร์ต่างๆ สำหรับให้เหมาะสมกับภาษาไทย เช่น การปรับค่า learning rate หรือจำนวนชั้นของ Transformer (Vaswani et al., 2017) (Accuracy Optimizing)
- การส่งออก Model (Export) หรือลดขนาด Model เพื่อให้ผู้ใช้งานตามบ้านสามารถนำไปติดตั้งได้โดยง่าย (Export & Quantization)

14.2 ตารางแผนงานวิจัย

แผนงานวิจัย	ผู้รับผิดชอบ	ช่วงระยะเวลาในการดำเนินงาน											
		1	2	3	4	5	6	7	8	9	10	11	12
1. เก็บข้อมูลและประมวลผลเพื่อให้เหมาะสมกับภาษาไทย	ดร. ปรัชญา บุญขวัญ												
2. พัฒนา/เทรน Pre-training Model LLM ภาษาไทย	ดร. ปรัชญา บุญขวัญ												
3. เตรียมข้อมูลในการตอบคำถาม (Instruction Dataset) สำหรับภาษาไทย	ดร. กอบก ฤตย์ วิริยะยุทธกร												
4. ปรับปรุงประสิทธิภาพของแบบจำลองด้วยการใช้ Transfer Learning / RLHF	ดร. สุธเมธ ยืนยง												
5. ประเมินประสิทธิภาพ	ทั้งหมด												
6. ปรับปรุงประสิทธิภาพโมเดล (Accuracy Optimization)	ดร. กอบก ฤตย์ วิริยะยุทธกร + ดร. สุธเมธ ยืนยง												

		4. LLM ภาษาไทย ที่พร้อมติดตั้งลงเครื่องคอมพิวเตอร์ผู้ใช้งานตามบ้าน	<ul style="list-style-type: none"> ○ สรุปความภาษาไทย ○ แต่งนิทาน ○ เขียนรายงาน ● เชื่อมต่อกับแหล่งความรู้ภายนอกได้ ● ทำข้อสอบ O-Net ได้ ● สามารถดาวน์โหลดและติดตั้งบนเครื่องคอมพิวเตอร์ได้ <ol style="list-style-type: none"> 2. เอกสารคู่มือการติดตั้ง 3. เอกสารคู่มือการใช้งาน 4. บทความวิชาการระดับนานาชาติ อย่างน้อย 1 ฉบับ
--	--	--	--

17.2 ผลงานส่งมอบ

ผลงานส่งมอบที่คาดว่าจะได้รับ	รายละเอียดของผลงานส่งมอบ	หน่วยนับ	สถานที่จัดเก็บและการใช้ประโยชน์ (กรณีมีต้นแบบ)
Model Thai LLM Opensource	<p>ความสามารถ</p> <ul style="list-style-type: none"> ● ตอบคำถามภาษาไทยในรูปแบบผู้ช่วยได้ ● ภาษาไทย/ภาษาอังกฤษ ● Few-shot Learning <ul style="list-style-type: none"> ○ แปลภาษาไทย/อังกฤษ ○ สรุปความภาษาไทย ○ แต่งนิทาน ○ เขียนรายงาน ● เชื่อมต่อกับแหล่งความรู้ภายนอกได้ ● ทำข้อสอบ O-Net ได้ ● สามารถดาวน์โหลดและติดตั้งบนเครื่องคอมพิวเตอร์ได้ <p>ส่งมอบ Source Code, Weight ของโมเดล</p>	1	เปิดเผยแพร่แบบ Opensource ให้ผู้ใช้งานสามารถ Download ได้ฟรี
Dataset ที่ใช้ในการพัฒนาทั้งหมด	<ul style="list-style-type: none"> - Dataset ที่ใช้ในการ Pretrain - Thai Instruct Dataset 	1	เปิดเผยแพร่แบบ Opensource ให้ผู้ใช้งานสามารถ Download ได้ฟรี
เอกสารคู่มือการติดตั้ง	<ul style="list-style-type: none"> - เอกสารคู่มือการติดตั้งการใช้งานบนเครื่องคอมพิวเตอร์ 	1	เปิดเผยแพร่แบบ Opensource ให้ผู้ใช้งานสามารถ Download ได้ฟรี
เอกสารคู่มือการใช้งาน	<ul style="list-style-type: none"> - เอกสารคู่มือการใช้งานบนเครื่องคอมพิวเตอร์ 	1	เปิดเผยแพร่แบบ Opensource ให้ผู้ใช้งานสามารถ Download ได้ฟรี

บทความวิชาการระดับ นานาชาติ อย่างน้อย 1 ฉบับ	บทความวิชาการระดับนานาชาติ อย่างน้อย PDF และ Source code ในรูปแบบ LaTeX	1	เปิดเผยแพร่แบบ Opensource ให้ผู้ใช้งานสามารถ Download ได้ฟรี
---	---	---	---

หมายเหตุ: ตีรายละเอียดคำจำกัดความของประเภทผลผลิต (ผลงานส่งมอบ) ผลลัพธ์และผลกระทบใน
ภาคผนวก

Impact Pathway

ปัจจัยนำเข้า (Input) ปัจจัยนำเข้าซึ่งเป็นปัจจัยที่ขับเคลื่อนให้ งานวิจัยดำเนินการสำเร็จและสร้างผล ประทบต่อสังคม	ผลผลิต (Output) ผลที่เกิดขึ้นสิ่งแรกและชัดเจนที่สุดจาก โครงการวิจัย โดยตอบวัตถุประสงค์การศึกษา ที่ตั้งไว้	ผลลัพธ์ (Outcome) คือ ผลจากการนำผลผลิตจากงานวิจัยไปใช้ประโยชน์โดยกลุ่มเป้าหมาย (User) ทำให้มีการเปลี่ยนแปลง (Change) พฤติกรรม การยอมรับด้านความรู้ ทักษะ และ ทักษะ	ผลกระทบ (Impact) คือ การเปลี่ยนแปลงจากผลลัพธ์ ในวงกว้าง
(1) งบประมาณการวิจัย (2) บุคลากร : นักวิจัย (3) องค์ความรู้เดิม หรือผลการศึกษา (Output) จากโครงการวิจัยก่อน หน้านี้ ที่ใช้ต่อยอดในการวิจัย (หากมี)		User หรือ ผู้ใช้ประโยชน์จากผลผลิต ของงานวิจัย งานวิจัยที่เกิดผลลัพธ์ ที่ สำคัญต้องมีผู้ใช้ประโยชน์ (User) มี การยอมรับ (Adoption) หรือการ นำไปใช้ในหลายระดับ 1. 1st User 2. 2nd User 3. Final User	การพิจารณาระดับความเปลี่ยนแปลง (Change) สามารถพิจารณาการใช้ ประโยชน์จากผลผลิตแยกตาม user แต่ ละประเภท เมื่อผู้ใช้ประโยชน์จาก งานวิจัยนำงานวิจัยไปใช้ประโยชน์อย่าง กว้างขวางในเชิงพาณิชย์ สามารถสร้าง รายได้สุทธิเพิ่มขึ้นจนทำให้เกิดการ เปลี่ยนแปลงในระดับรายได้สุทธิ และ คุณภาพชีวิตที่ดีขึ้น
(1) งบประมาณงานวิจัย 10 ล้านบาท	(1) Opensource Thai LLMs ที่ตอบคำถามให้กับ Business และ SME ใน ไทยได้	(1) Thai Tech Startup สามารถสร้าง AI Generative Startup ได้ ต่อยอด พัฒนาผลิตภัณฑ์ ใหม่ๆ ขึ้นมาได้	(1) SME, Business สามารถสร้างรายได้สุทธิ เพิ่มขึ้นโดยใช้ AI เข้ามา ลดต้นทุนการผลิต และลด ค่าใช้จ่ายในการติดต่อ ประสานงานลูกค้า (1) ผลกระทบทาง เศรษฐกิจ เพิ่มขึ้นโดยตรง500 ล้าน บาทในปี 2023 และจะ เป็น 2,000 ล้านบาทในปี 2030 เพิ่มขึ้นโดยอ้อมจะ Productivity ที่สูงขึ้น

				<p>คาดว่าจะช่วย GDP ประเทศไทยให้ได้สูงขึ้นอย่างน้อย 5% => 0.5 ล้านล้านบาท</p>
<p>(2) นักวิจัย: 3 คน ผู้ช่วยวิจัย: 3 คน ช่างเทคนิค: 5 คน</p>	<p>(2) เจ้าหน้าที่ผู้เชี่ยวชาญด้าน Generative AI (LLM) ของไทยจำนวน 11 ท่าน</p>	<p>(2) Users ผู้ใช้งานระบบจะได้ AI Assistant ที่มีความฉลาด และใช้งานฟรี</p>	<p>(2) Users ผู้ใช้งานระบบจะได้ AI Assistant ที่มีความฉลาด และใช้งานฟรี และช่วยงาน เพิ่ม Productivity ได้ คุณภาพชีวิตที่ดีขึ้น</p>	<p>(2) ผลกระทบทางสังคม สกิลการทำงานของ White Collar ที่เพิ่มขึ้นจากความช่วยเหลือของ Thai LLM ทำให้คนไทยพร้อมกับการเปลี่ยนแปลงของยุคปฎิวัติปัญญาประดิษฐ์</p>
<p>(3) Thai Language Dataset อย่างน้อย 500GB และ Instruct Dataset อย่างน้อย 80K คู่</p>	<p>(4) Dataset เปิดเผยต่อสาธารณะสามารถให้ Startup SME นำไปต่อยอดได้</p>	<p>(3) Business / SME ธุรกิจจะได้ใช้ประโยชน์จากปัญญาประดิษฐ์ สามารถช่วยประมวลผล และตอบคำถามลูกค้าได้อย่างรวดเร็ว สามารถสร้าง Productivity และลดค่าใช้จ่ายได้อย่าง</p>	<p>(3) Thai Tech Startup มีอาวุธในการออกแบบผลิตภัณฑ์ด้านปัญญาประดิษฐ์ใหม่ๆ พร้อมกับการเป็นผู้นำใน Southeast Asian</p>	

		มหาศาล เป็นผู้นำด้าน AI ในอาเซียน		
(4) เครื่องประมวลผล ขนาดใหญ่ (GPU Super Computer)				

Thai LLM Opensource Business Model



<http://www.businessmodelgeneration.com>
Designed by: Business Model Foundry AG
The makers of Business Model Generation and Strategyzer

การบริหารความเสี่ยง (Risk Management)

19.1 ความเสี่ยงจากปัจจัยภายนอก เช่น การสั่งซื้ออุปกรณ์ ครุภัณฑ์ วัตถุดิบตามฤดูกาล ฯลฯ ให้อธิบายมาตรการหรือแนวทางในการป้องกัน แก้ไข บรรเทาความเสี่ยงที่อาจจะเกิดขึ้น

- เครื่อง Super Computer ขนาดใหญ่ที่ใช้ในงานวิจัยนี้ อาจจะประสิทธิภาพไม่พร้อม ต่อการใช้งานเพื่อการพัฒนา
- ทำการสำรวจผู้ให้บริการ Super Computer ขนาดใหญ่ และทำการศึกษาให้ดีกว่าก่อนที่จะตัดสินใจใช้บริการ

19.2 ความเสี่ยงจากปัจจัยภายใน เช่น การใช้เครื่องมือส่วนรวม ฯลฯ ให้อธิบายมาตรการหรือแนวทางในการป้องกัน แก้ไข บรรเทาความเสี่ยงที่อาจจะเกิดขึ้น

- การ Train โมเดลขนาดใหญ่แล้วไม่สามารถนำมาใช้งานได้ เนื่องจากค่า Configuration ที่ผิดพลาด
- แก้ไข: ทดลองกับโมเดลเล็ก จำนวน Epoch น้อยๆ เพื่อทดสอบเสียก่อน

ข้าพเจ้าขอรับรองว่า ข้อความในข้อเสนอโครงการนี้ ไม่มีการคัดลอกเนื้อหาจากแหล่งข้อมูลอื่น และยินยอมให้มีการนำข้อเสนอโครงการพร้อมข้อมูลทั้งหมด ไปใช้ในการประเมิน และพิจารณากลับกรองโครงการ

ลายมือชื่อ.....
(ชื่อหัวหน้าโครงการ)

ภาคผนวก คำอธิบาย

Technology Readiness Level – TRL

Technology Readiness Level – TRL คือ การบ่งชี้ระดับความพร้อมและเสถียรภาพของเทคโนโลยีตามบริบทการใช้งาน ตั้งแต่วัตถุดิบองค์ประกอบสำคัญ อุปกรณ์ และกระบวนการทำงานทั้งระบบ ก่อนที่จะมีการบูรณาการเทคโนโลยีเป็นระบบ รายละเอียดดังนี้

- TRL 1 – หลักการพื้นฐานได้รับการพิจารณาและมีการรายงาน
- TRL 2 – มีการสร้างรูปแบบหลักการ และ/หรือ การประยุกต์ใช้
- TRL 3 – หลักการได้ถูกสาธิตด้วยการวิเคราะห์หรือการทดลอง
- TRL 4 – องค์ประกอบที่สำคัญได้ถูกสาธิตในระดับห้องปฏิบัติการแล้ว
- TRL 5 – องค์ประกอบที่สำคัญได้ถูกสาธิตในระดับสถานะที่ใกล้เคียงกับการใช้งาน
- TRL 6 – ตัวแทนสิ่งที่จะส่งมอบได้ถูกสาธิตในระดับสถานะที่ใกล้เคียงกับการใช้งาน
- TRL 7 – ผลของการพัฒนาขั้นสุดท้ายได้ถูกสาธิตในสถานะทำงาน
- TRL 8 – เทคโนโลยีที่ส่งมอบ ได้ผ่านการทดสอบและสาธิตในสภาพการใช้งานจริง
- TRL 9 – เทคโนโลยีที่ส่งมอบได้ผ่านการใช้งานจริง

Societal Readiness Level – SRL

Societal Readiness Level – SRL คือ ระดับความพร้อมของความรู้และเทคโนโลยีทางด้านสังคม ที่ใช้ในการประเมินระดับความพร้อมของความรู้และเทคโนโลยีทางด้านสังคม องค์ความรู้ เทคโนโลยี กระบวนการ การแก้ปัญหา สิ่งประดิษฐ์และนวัตกรรมทั้งด้านสังคม เป็นเครื่องมือที่นำมาประยุกต์ใช้เพื่อสร้างความเข้าใจร่วมกัน ในการบริหารจัดการโครงการ โปรแกรมทางด้านสังคม รายละเอียดดังนี้

- SRL 1 – การวิเคราะห์ปัญหาและกำหนดความพร้อมของความรู้และเทคโนโลยีทางด้านสังคมที่มี (identifying problem and identifying societal readiness)
- SRL 2 – การกำหนดปัญหา การเสนอแนวคิดในการพัฒนาหรือการแก้ปัญหาและคาดการณ์ผลกระทบที่อาจเกิดขึ้น และระบุผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้องในโครงการ (formulation of problem, proposed solution(s) and potential impact, expected societal readiness; identifying relevant stakeholders for the project.)
- SRL 3 – ศึกษา วิจัย ทดสอบแนวทางการพัฒนาหรือแก้ปัญหาที่กำหนดขึ้นร่วมกับผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้อง (initial testing of proposed solution(s) together with relevant stakeholders)
- SRL 4 – ตรวจสอบแนวทางการแก้ปัญหาโดยการทดสอบในพื้นที่นำร่องเพื่อยืนยันผลกระทบตามที่คาดว่าจะเกิดขึ้น และดูความพร้อมขององค์ความรู้และเทคโนโลยี (problem validated through pilot testing in relevant environment to substantiate proposed impact and societal readiness)
- SRL 5 – แนวทางการแก้ปัญหาได้รับการตรวจสอบ ถูกนำเสนอแก่ผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้อง area (proposed solution (s) validated, now by relevant stakeholders in the area)
- SRL 6 – ผลการศึกษานำไปประยุกต์ใช้ในสิ่งแวดล้อมอื่น และดำเนินการกับผู้มีส่วนได้ส่วนเสียที่เกี่ยวข้องเพื่อให้ได้ข้อเสนอแนะเบื้องต้นเพื่อให้เกิดผลกระทบที่เป็นไปได้ (solution (s) demonstrated in relevant environment and in cooperation with relevant stakeholders to gain initial feedback on potential impact)
- SRL 7 – การปรับปรุงโครงการและ/หรือการแนวทางการพัฒนา การแก้ปัญหา รวมถึงการทดสอบแนวทางการพัฒนา การแก้ปัญหาใหม่ในสภาพแวดล้อมที่เกี่ยวข้องกับผู้มีส่วนได้ส่วนเสีย (refinement of project and/or solution and, if needed, retesting in relevant environment with relevant stakeholders)

- SRL 8 – เสนอแนวทางการพัฒนา การแก้ปัญหาในรูปแบบแผนการดำเนินงานที่สมบูรณ์ และได้รับการยอมรับ (proposed solution(s) as well as a plan for societal adaptation complete and qualified)
- SRL 9 – แนวทางการพัฒนาและการแก้ปัญหาของโครงการได้รับการยอมรับและสามารถนำไปประยุกต์ใช้ได้กับสิ่งแวดล้อมอื่น ๆ (actual project solution (s) proven in relevant environment)

ตัวอย่าง

Business model canvas

Name: Live Shop

Key partners <small>Who are your most important partners? Which key resources do you acquire from partners? Which key activities do your partners perform?</small> <ul style="list-style-type: none"> Fashion Events e.g. Zaap on sale Beauty influencers (e.g. Pimtha) Payment Providers Logistic companies 	Key activities <small>What are the activities you perform every day to create & deliver your value proposition?</small> <ul style="list-style-type: none"> Marketing to shoppers and business owners Research into scaling opportunities in the future Key resources <small>What are the resources you need to create & deliver your value proposition?</small> <ul style="list-style-type: none"> Developers Database Marketers Designers 	Value propositions <small>What is the value you deliver to your customer? Which of your customer's problems are you helping to solve? What is the customer need that your value proposition addresses? What is your promise to your customers? What are the products and services you create for your customers?</small> <ul style="list-style-type: none"> Try-on and buy clothes online (Mix & Match) Visualize looks attractive for your body shape Personalized shopping Reach out to a wider consumer base Provide insightful information for the business owner Saves time to answer (line) chat about sizes 	Customer relationships <small>What relationship does each customer segment expect you to establish and maintain?</small> <ul style="list-style-type: none"> <input type="checkbox"/> Multimedia marketing (e.g. beauty influencers on FB & IG, Pop-up Booth, YouTube, 6-second ad.) <input type="checkbox"/> Direct marketing to Business owners Channels <small>How does your value proposition reach your customer? Where can your customer buy or use your products or services?</small> <ul style="list-style-type: none"> Mobile application Instagram, Line, Facebook, Twitter 	Customer segments <small>For whom are you creating value? What are the customer segments that expect pay, receive or decide on your value proposition?</small> <ul style="list-style-type: none"> Thai millennials in Bangkok Thai fashion brand owners
Cost structure <small>What are the important costs you must incur to create & deliver your value proposition?</small> <ul style="list-style-type: none"> Fixed: Technology development, R & D costs Variable: Marketing Costs, Team staff wages, Partnership fees 		Revenue streams <small>How do customers reward you for the value you provide to them? What are the different revenue models?</small> <ul style="list-style-type: none"> <input type="checkbox"/> Commission from marketing <input type="checkbox"/> Commission from sales (Phase 2) <input type="checkbox"/> Monthly Subscription to date report <input type="checkbox"/> Integration with big companies' websites 		

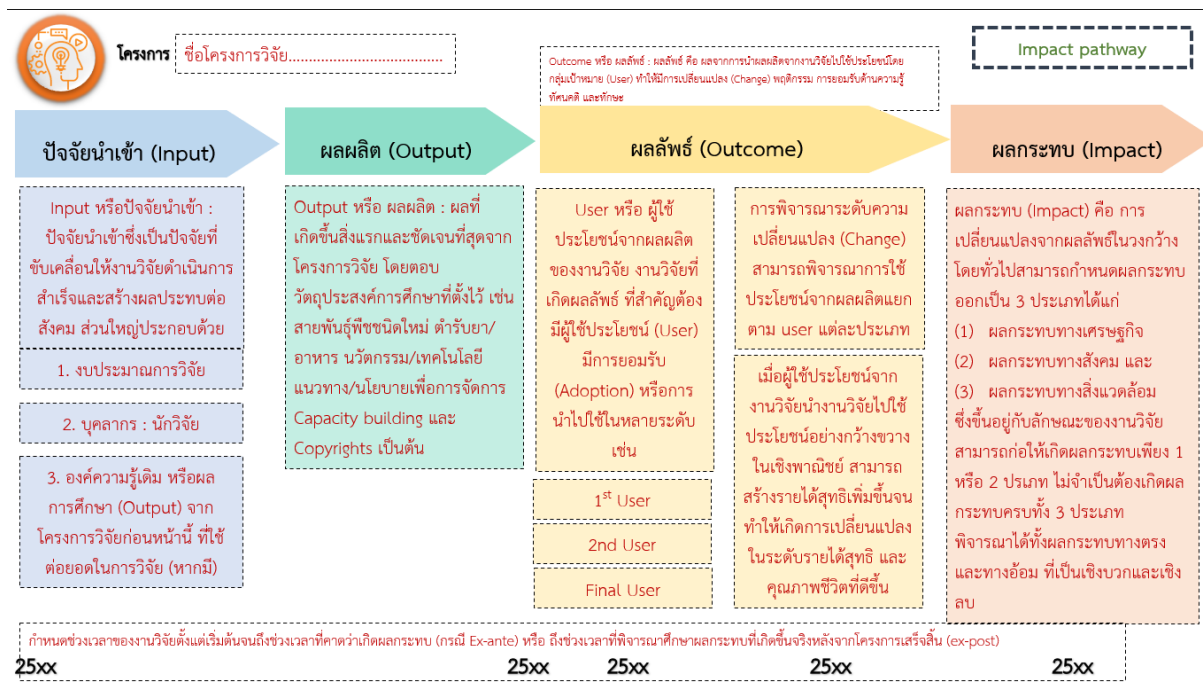
Brought to you by Business Models Inc.

www.strategyzer.com

Technology Evaluation Canvas

Identifying MVP		Value Perception		Differentiate		Interviews
Features	Importance for Customer (+3 & -3)	Customer Need	Customer Benefit	Technology (+3 & -3)	Competitor	Technical
Tech to Market		Problem Validation		Problem-Solution Fit		Solution - Market Fit
TRL Level / Development Levels		Problem		Solution		User
Co-Innovators/ Partners		Value Proposition		Customer Segment		Decision Maker
		IPR Status		Risk and Barriers		Partners

คำอธิบาย Impact pathway



ประเภทของผลผลิตและคำจำกัดความ (Type of Outputs and Definition)

(อ้างอิงจากคู่มือการจัดทำคำของบประมาณและการบริหารจัดการงบประมาณเพื่อสนับสนุนงานมูลฐานตามพันธกิจของหน่วยรับงบประมาณ (Fundamental Fund; FF) ประจำปีงบประมาณ พ.ศ. 2566)

1. นิยามของผลผลิต คือ ผลผลิตที่เกิดขึ้นจากการวิจัยที่ได้รับการจัดสรรทุนวิจัย ผ่านกองทุนส่งเสริมวิทยาศาสตร์ วิจัย และนวัตกรรม โดยเป็นผลที่เกิดขึ้นทันทีเมื่อจบโครงการ และเป็นผลโดยตรงจากการดำเนินโครงการ ทั้งนี้ หน่วยงานจะต้องนำเสนอภายใน 2 ปีงบประมาณ
2. ประเภทของผลผลิต ประกอบด้วย 10 ผลผลิต ตามตารางดังนี้

ประเภทของผลผลิต (Type of Outputs)	คำจำกัดความ (Definition)
1. กำลังคน หรือหน่วยงาน ที่ได้รับการพัฒนาทักษะ	กำลังคนหรือหน่วยงานเป้าหมายที่ได้รับการพัฒนาจากโครงการ ววน. โดยนับเฉพาะคนหรือ หน่วยงานที่เป็นเป้าหมายของโครงการนั้น ๆ ซึ่งอาจเป็นโครงการในรูปแบบทุนการศึกษา การฝึกอบรมเพื่อเพิ่มทักษะ หรือการดำเนินการในรูปแบบอื่นที่ระบุไว้ในโครงการ
2. ต้นฉบับบทความวิจัย (Manuscript)	งานเขียนทางวิชาการ ซึ่งมีการกำหนดประเด็นที่ต้องการอธิบายหรือวิเคราะห์อย่างชัดเจน ทั้งนี้ ต้องมีการวิเคราะห์ประเด็นดังกล่าวตามหลักวิชาการ โดยมีการสำรวจวรรณกรรมเพื่อสนับสนุน จนสามารถสรุปผลการวิเคราะห์ในประเด็นนั้นได้ มีการแสดงเหตุผลหรือที่มาของประเด็นที่ต้องการอธิบายหรือวิเคราะห์ กระบวนการอธิบายและวิเคราะห์และบทสรุป มีการอ้างอิงและบรรณานุกรมที่ครบถ้วนและสมบูรณ์วารสารการวิจัยนั้นอาจจะเผยแพร่เป็นรูปเล่มสิ่งพิมพ์หรือ เป็นสื่ออิเล็กทรอนิกส์ ซึ่งต้นฉบับบทความวิจัย (Manuscript) ได้แก่ Proceeding ระดับชาติ, Proceeding ระดับนานาชาติ, บทความในประเทศ และบทความต่างประเทศ
3. หนังสือ	ข้อมูลงานวิจัยในรูปแบบหนังสือ ตำรา หรือหนังสืออิเล็กทรอนิกส์ (E-book) ทั้งระดับชาติและ นานาชาติ โดยจะต้องผ่านกระบวนการ Peer review ประกอบด้วย 3.1 บางบทของหนังสือ (Book Chapter) 3.2 หนังสือทั้งเล่ม (Whole book) 3.3 เอกสาร/หนังสือที่มีเนื้อหาเกี่ยวกับเรื่องใดเรื่องหนึ่ง อย่างครบถ้วน (Monograph)
4. ต้นแบบผลิตภัณฑ์ หรือ เทคโนโลยี/กระบวนการใหม่ หรือนวัตกรรมทางสังคม	ผลงานที่เกิดจากการวิจัยและพัฒนานวัตกรรม ที่ทำให้เกิดผลิตภัณฑ์ใหม่ หรือเทคโนโลยีใหม่/ กระบวนการใหม่ หรือการปรับปรุงผลิตภัณฑ์ หรือเทคโนโลยี/ กระบวนการให้ดีขึ้นกว่าเดิม รวมถึงสื่อสร้างสรรค์ สื่อสารคดีเพื่อการเผยแพร่ สื่อออนไลน์ แอปพลิเคชัน / Podcast / กิจกรรม / กระบวนการ เพื่อสร้างการเรียนรู้ การมีส่วนร่วม และ/หรือ การตระหนักรู้ต่าง ๆ 4.1 ต้นแบบผลิตภัณฑ์ หมายถึง ต้นแบบในรูปแบบของผลิตภัณฑ์ที่ใช้สำหรับการทดสอบก่อน สิ่งผลิตจริง ที่พัฒนาขึ้นจากกระบวนการวิจัย พัฒนา หรือการปรับปรุงกระบวนการเดิมด้วยองค์ ความรู้ด้านวิทยาศาสตร์และเทคโนโลยีทั้งในระดับห้องปฏิบัติการ ระดับภาคสนาม ระดับอุตสาหกรรม

ประเภทของผลผลิต (Type of Outputs)	คำจำกัดความ (Definition)
	<p>4.2 เทคโนโลยี/กระบวนการใหม่ หมายถึง กรรมวิธีขั้นตอน หรือเทคนิค ที่พัฒนาขึ้นจากกระบวนการวิจัย พัฒนา หรือการปรับปรุงกระบวนการเดิมด้วยองค์ความรู้ด้านวิทยาศาสตร์ และเทคโนโลยี</p> <p>4.3 นวัตกรรมทางสังคม (Social Innovation) หมายถึง การประยุกต์ใช้ความคิดใหม่ และ เทคโนโลยีที่เหมาะสม ในการยกระดับคุณภาพชีวิต ชุมชน และสิ่งแวดล้อม อันจะนำไปสู่ความเท่าเทียมกันในสังคม และสามารถลดปัญหาความเหลื่อมล้ำได้อย่างเป็นรูปธรรม ตัวอย่างเช่น หลักสูตรอบรมปฏิบัติการเพื่อพัฒนานักวิจัย, หลักสูตรพื้นฐานเพื่อพัฒนาอาชีพใหม่ในรูปแบบ Reskill หรือ Upskill) หลักสูตรการเรียนการสอน, หลักสูตรบัณฑิตพันธุ์ใหม่ หลักสูตรการผลิต ครู เป็นต้น</p>
5. ทรัพย์สินทางปัญญา	ผลงานอันเกิดจากการประดิษฐ์ คิดค้น หรือสร้างสรรค์ของนักวิจัย ได้แก่ อนุสิทธิบัตร สิทธิบัตร การประดิษฐ์ สิทธิบัตรการออกแบบผลิตภัณฑ์ ลิขสิทธิ์ เครื่องหมายทางการค้า ความลับทางการค้า ชื่อทางการค้า การขึ้นทะเบียนพันธุ์พืชหรือสัตว์ สิ่งบ่งชี้ทางภูมิศาสตร์ แบบผังภูมิ ของวงจร
6. เครื่องมือ และโครงสร้างพื้นฐาน (Facilities and Infrastructure)	เครื่องมือ และโครงสร้างพื้นฐานเพื่อการพัฒนาวิทยาศาสตร์ เทคโนโลยี และการวิจัยและพัฒนา นวัตกรรม ที่จัดซื้อ สร้างขึ้น หรือพัฒนาต่อยอดภายใต้โครงการ
7. ฐานข้อมูล ระบบและกลไก หรือมาตรฐาน	<p>การพัฒนาฐานข้อมูล และสร้างระบบ กลไก หรือมาตรฐาน ที่ตอบสนองการพัฒนาวิทยาศาสตร์ วิจัยและนวัตกรรมต่าง ๆ และเอื้อต่อการพัฒนาประเทศ ไม่ว่าจะเป็นการพัฒนากำลังคน การจัดการปัญหาทางสังคม สิ่งแวดล้อม และการสร้างความสามารถในการแข่งขัน เป็นต้น</p> <p>- ระบบและกลไก หมายถึง ขั้นตอนหรือเครื่องมือ การปฏิบัติงานที่มีการกำหนดอย่างชัดเจนใน การดำเนินการ เพื่อให้ได้ผลออกมาตามที่ต้องการ ขั้นตอนการปฏิบัติงานจะต้องปรากฏให้ทราบ โดยทั่วกัน ไม่ว่าจะอยู่ในรูปของ เอกสาร หรือสื่ออิเล็กทรอนิกส์ หรือโดยวิธีการอื่น ๆ องค์ประกอบของระบบและกลไก ได้แก่ ปกติ นำเข้า กระบวนการ ผลผลิต กลุ่มคนที่เกี่ยวข้อง และข้อมูลป้อนกลับ ซึ่งมีความสัมพันธ์เชื่อมโยงกัน ตัวอย่างเช่น ระบบการผลิตและการพัฒนากำลังคน, ระบบส่งเสริมการจัดการทรัพยากรและ สิ่งแวดล้อม,ระบบส่งเสริมการวิจัยร่วมกับภาคอุตสาหกรรม, ระบบบริการหรือสิ่งสนับสนุน ประชาชนทั่วไป, ระบบบริการหรือสิ่งสนับสนุนกลุ่มผู้สูงอายุ, ระบบบริการหรือสิ่งสนับสนุนกลุ่ม ผู้ด้อยโอกาสรวมถึงกลไกการพัฒนาเชิงพื้นที่</p> <p>- ฐานข้อมูล (Database) คือ ชุดของสารสนเทศ ที่มีโครงสร้างสม่ำเสมอ หรือชุดของ สารสนเทศใด ๆ ที่ประมวลผลด้วยคอมพิวเตอร์ หรือสามารถประมวลด้วยคอมพิวเตอร์ได้</p> <p>- มาตรฐาน หมายถึง การรับรองมาตรฐานสินค้า และ/หรือ ศูนย์ทดสอบต่าง ๆ เพื่อสร้างและ ยกระดับความสามารถทางด้านคุณภาพ ทั้งในชาติและนานาชาติ</p>

ประเภทของผลผลิต (Type of Outputs)	คำจำกัดความ (Definition)
8. เครือข่าย	เครือข่ายความร่วมมือ (Network) และสมาคม (Consortium) ด้านวิทยาศาสตร์ วิจัยและ นวัตกรรม ที่เกิดจากการดำเนินการของโครงการ ทั้งเครือข่ายในประเทศ และเครือข่ายระดับ นานาชาติ ซึ่งจะช่วยในการยกระดับความสามารถในการ แข่งขันให้แก่ประเทศ ได้แก่ <ol style="list-style-type: none"> 1. เครือข่ายความร่วมมือทางด้านวิชาการ 2. เครือข่ายเพื่อการพัฒนาเศรษฐกิจ 3. เครือข่ายเพื่อการพัฒนาสังคม 4. เครือข่ายเพื่อการพัฒนาสิ่งแวดล้อม เป็นต้น
9. การลงทุนวิจัยและนวัตกรรม	ความสามารถในการระดมทุนเงินงบประมาณจากภาครัฐ และผู้ประกอบการ ภาคเอกชน ทั้งใน ประเทศและต่างประเทศ เพื่อการลงทุนสนับสนุนการวิจัยและ นวัตกรรม ทั้งในรูปของเงินสด (In cash) และส่วนสนับสนุนอื่นที่ไม่ใช่เงินสด (In kind)
10. ข้อเสนอแนะเชิงนโยบาย (Policy Recommendation) และมาตรการ (Measures)	ข้อเสนอแนะในระดับนโยบายและระดับปฏิบัติ หรือมาตรการจากงานวิจัยที่เป็น ประโยชน์ต่อ ภาคประชาชน สังคม และเศรษฐกิจ รวมทั้งสามารถนำไปใช้ในการ บริหารจัดการ และแก้ปัญหา ของประเทศ เช่น มาตรการที่ใช้เพื่อปรับปรุง กฎหมาย/ระเบียบ หรือพัฒนามาตรการและสร้าง แรงจูงใจให้เอื้อต่อการพัฒนาภาค ประชาชน สังคม หรือเศรษฐกิจ

ประเภทของผลลัพธ์และคำจำกัดความ (Type of Outcomes and Definition)

ประเภทของผลลัพธ์ (Types of Outcomes)	คำจำกัดความ (Definition)
ผลงานตีพิมพ์ (Publications)	ผลงานทางวิชาการในรูปแบบสิ่งพิมพ์และไฟล์อิเล็กทรอนิกส์ ซึ่งเกิดจากการ ศึกษาวิจัย อาทิเช่น บทความจากการประชุมวิชาการ บทความวิจัย บทความ ปริทัศน์ บทความวิชาการ หนังสือ ตำรา พจนานุกรม และงานวิชาการอื่นๆ ใน ลักษณะเดียวกัน
การอ้างอิง (Citations)	จำนวนครั้งในการอ้างอิงผลงานวิจัยที่ตีพิมพ์ในวารสารระดับนานาชาติ โดยสืบค้น จากฐานข้อมูล Scopus
เครื่องมือและระเบียบวิธีการวิจัย (Research tools and methods)	เครื่องมือหรือกระบวนการที่ผู้วิจัยใช้ในการทดลอง ทดสอบ เก็บรวบรวมหรือ วิเคราะห์ข้อมูล โดยเป็นสิ่งใหม่ที่ไม่ได้มีมาก่อน แต่ได้เผยแพร่และเป็นที่ยอมรับโดย มีผู้นำเครื่องมือและระเบียบวิธีการวิจัยไปใช้ต่อและมีหลักฐานอ้างอิงได้
ฐานข้อมูลและแบบจำลองวิจัย (Research databases and models)	ฐานข้อมูล (ระบบที่รวบรวมข้อมูลไว้ในที่เดียวกัน) หรือแบบจำลอง (การสร้าง รูปแบบเพื่อแทนวัตถุ กระบวนการ ความสัมพันธ์ หรือ สถานการณ์) ที่ถูกพัฒนาขึ้น จากงานวิจัย โดยมีผู้นำฐานข้อมูลหรือแบบจำลองไปใช้ให้เกิดประโยชน์มีหลักฐาน อ้างอิงได้

ประเภทของผลลัพธ์ (Types of Outcomes)	คำจำกัดความ (Definition)
ความก้าวหน้าในวิชาชีพของบุคลากร ด้านวิทยาศาสตร์วิจัยและนวัตกรรม (Next destination)	การติดตามการเคลื่อนย้ายและความก้าวหน้าในวิชาชีพของบุคลากรในโครงการ ด้านวิทยาศาสตร์ วิจัยและนวัตกรรม (ววน.) ที่ได้รับงบประมาณสนับสนุนจาก กองทุนส่งเสริม ววน. หลังจากสิ้นสุดโครงการ โดยมีหลักฐานอ้างอิงได้
รางวัลและการยอมรับ (Awards and recognition)	เกียรติยศ รางวัลและการยอมรับจากสังคมที่ได้มาโดยหน้าที่การทำงานจากการ ทำงานด้านวิทยาศาสตร์ วิจัยและนวัตกรรม (ววน.) โดยมีส่วนที่เกี่ยวข้องกับ โครงการที่ได้รับงบประมาณจากกองทุนส่งเสริม ววน. โดยมีหลักฐานอ้างอิงได้
การใช้ประโยชน์จากเครื่องมือ อุปกรณ์ ห้องวิจัยและโครงสร้างพื้นฐาน (Use of facilities and resources)	การใช้ประโยชน์จากเครื่องมือ อุปกรณ์ ห้องวิจัยและโครงสร้างพื้นฐานด้าน วิทยาศาสตร์ วิจัยและนวัตกรรม (ววน.) ที่นักวิจัยพัฒนาขึ้น หรือได้รับงบประมาณ เพื่อการจัดทำให้เกิดประโยชน์ต่อผู้ที่มาใช้งานในวงกว้าง โดยมีหลักฐานอ้างอิงได้
ทรัพย์สินทางปัญญาและการอนุญาตให้ ใช้สิทธิ (Intellectual property and licensing)	ทรัพย์สินทางปัญญา หมายถึง การประดิษฐ์ คิดค้นหรือคิดทำขึ้น อันเป็นผลใด มาซึ่งผลิตภัณฑ์หรือกรรมวิธีใดชิ้นใหม่ หรือการกระทำใดๆ ที่ทำให้ดีขึ้นซึ่ง ผลิตภัณฑ์หรือกรรมวิธี หรือการกระทำใดๆ เกี่ยวกับงานที่ผู้สร้างสรรค์ได้ริเริ่มโดย ใช้สติปัญญาความรู้ ความสามารถ และความวิริยะอุตสาหะของตนเองในการสร้าง ให้เกิดงานสร้างสรรค์ 9 ประเภทตามที่กฎหมายกำหนด อาทิเช่น งานวรรณกรรม งานศิลปกรรม งานดนตรีกรรม งานภาพยนตร์ เป็นต้น โดยไม่ลอกเลียนงานของ ผู้อื่น ซึ่งเกิดจากผลงานวิจัยด้านวิทยาศาสตร์ วิจัยและนวัตกรรม (ววน.) ที่ได้รับ งบประมาณสนับสนุนจากกองทุนส่งเสริม ววน. โดยมีหลักฐานอ้างอิงได้ การอนุญาตให้ใช้สิทธิ หมายถึง การที่เจ้าของสิทธิอนุญาตให้ผู้ขอใช้สิทธิใดๆ ที่ เกิดขึ้นจากงานวิจัย เช่น ผลิต / ขาย / ใช้ หรือมีไว้ โดยไม่มีการเปลี่ยนแปลงความ เป็นเจ้าของสิทธิทั้งนี้เพื่อประโยชน์เชิงพาณิชย์เป็นหลัก โดยมีหลักฐานอ้างอิงได้
การจัดตั้งบริษัท (Spin-off Companies)	การนำเอาเทคโนโลยีหรือองค์ความรู้ใหม่หาวิทยาลัยที่เกิดจากการวิจัย (technology transfer) มาจัดตั้งเป็นบริษัท เพื่อขับเคลื่อนงานวิจัยไปสู่การขยาย ผลในเชิงพาณิชย์ โดยมีหลักฐานอ้างอิงได้
ผลิตภัณฑ์ใหม่ (New Products)	ผลิตภัณฑ์ใหม่ หมายถึง ผลิตภัณฑ์ประเภทต่างๆ ที่ได้จากการวิจัย อาทิเช่น ผลิตภัณฑ์ทางการแพทย์ / ผลิตภัณฑ์ที่เกี่ยวข้องกับซอฟต์แวร์และ ปัญญาประดิษฐ์ / ผลิตภัณฑ์ด้านเทคนิคและเทคโนโลยี / ผลิตภัณฑ์ทางการแพทย์ และอาหาร และผลิตภัณฑ์ด้านศิลปะและการสร้างสรรค์ โดยผลิตภัณฑ์ด้านการ สร้างสรรค์ หมายถึงถึงผลิตภัณฑ์ทางด้านการท่องเที่ยว เช่น เส้นทาง ท่องเที่ยว การจัดโปรแกรมด้านการท่องเที่ยวเพื่อนำไปสู่รูปแบบการท่องเที่ยว แบบใหม่ๆ เป็นต้น โดยเป็นสิ่งที่ถูกคิดค้น พัฒนาขึ้นอย่างมีนัยยะสำคัญในโครงการ และสามารถก่อให้เกิดมูลค่าทางเศรษฐกิจ โดยมีหลักฐานอ้างอิงได้
ทุนต่อยอด (Further funding)	ทุนที่นักวิจัยได้รับเงินอุดหนุนการวิจัยต่อยอดจากงานวิจัยเดิม ซึ่งเกิดจากการนำ ผลงานวิจัยที่ได้ของโครงการวิจัยเดิมมาเขียนเป็นข้อเสนอโครงการเพื่อขอรับทุน

ประเภทของผลลัพธ์ (Types of Outcomes)	คำจำกัดความ (Definition)
	วิจัยต่อยอดในโครงการใหม่ สิ่งสำคัญคือ การให้ข้อมูลเกี่ยวกับแหล่งทุนและงบประมาณที่ได้รับจากโครงการทุนวิจัยต่อยอดใหม่ โดยมีหลักฐานอ้างอิงได้
ความร่วมมือหรือหุ้นส่วนความร่วมมือ (Collaborations and partnerships)	ความร่วมมือหรือหุ้นส่วนความร่วมมือที่เกิดขึ้นหลังจากโครงการวิจัยเสร็จสิ้น โดยเป็นความร่วมมือที่เกี่ยวข้องโดยตรงหรืออาจจะทางอ้อมจากการดำเนินโครงการ ทั้งนี้สิ่งสำคัญคือ การระบุผลผลิต (output) ผลลัพธ์ (outcome) และผลกระทบ (impact) ที่เกิดขึ้นจากความร่วมมือหรือหุ้นส่วนความร่วมมือนี้ โดยมีหลักฐานอ้างอิงได้
นโยบาย แนวปฏิบัติ แผนและกฎระเบียบ (Policy, practice, plan and regulations)	การดำเนินการอย่างใดอย่างหนึ่งเพื่อขับเคลื่อนการนำผลงานวิจัยไปสู่การใช้ประโยชน์เชิงนโยบาย หรือเกิดแนวปฏิบัติ แผนและกฎระเบียบต่างๆ ขึ้นใหม่ เพื่อให้เกิดผลลัพธ์และผลกระทบต่อการเปลี่ยนแปลงในมิติต่างๆ ทางเศรษฐกิจ สังคมวัฒนธรรม สิ่งแวดล้อม และการเมืองการปกครอง ที่เป็นประโยชน์ต่อสังคมและประเทศโดยรวม โดยมีหลักฐานอ้างอิงได้ ทั้งนี้ต้องไม่ใช่การดำเนินการที่ระบุไว้เป็นส่วนหนึ่งของแผนงานวิจัย
กิจกรรมสร้างการมีส่วนร่วม (Engagement activities)	กิจกรรมที่หัวหน้าโครงการและ/หรือทีมวิจัย ได้สื่อสารผลงานด้านวิทยาศาสตร์ วิจัยและนวัตกรรม (ววน.) กับกลุ่มผู้เข้าร่วมกิจกรรม เพื่อขับเคลื่อนให้เกิดการนำผลงานวิจัยไปใช้ประโยชน์ และเป็นเส้นทางที่ส่งผลให้เกิดผลกระทบในวงกว้างต่อไป โดยมีหลักฐานอ้างอิงได้ ทั้งนี้กิจกรรมดังกล่าวต้องมีใช้กิจกรรมที่ได้ระบุไว้เป็นส่วนหนึ่งของแผนงานวิจัย

ผลกระทบที่คาดว่าจะเกิดขึ้น (Expected Impact)

นิยามของผลกระทบ คือ การเปลี่ยนแปลงที่เกิดขึ้นจากผลลัพธ์ (outcome) ในวงกว้างทั้งด้านวิชาการ นโยบาย เศรษฐกิจ สังคมและสิ่งแวดล้อม หรือผลสำเร็จระยะยาวที่เกิดขึ้นต่อเนื่องจากการเปลี่ยนแปลงของผลลัพธ์ โดยผ่านกระบวนการการสร้างการมีส่วนร่วม (Engagement activities) และมีเส้นทางของผลกระทบ (impact pathway) ในการขับเคลื่อนไปสู่การสร้างผลกระทบ ทั้งนี้ ผลกระทบที่เกิดขึ้นจะพิจารณารวมผลกระทบในเชิงบวกและเชิงลบ ทางตรงและทางอ้อม ทั้งที่ตั้งใจและไม่ตั้งใจให้เกิดขึ้น