**ที่มาและความสำคัญ**

เนื่องจากคนเรารับรู้ภาพได้ดีกว่าข้อข้อความจึงเกิดการนำเสนอแบบภาพขึ้นมา ในปัจจุบันเราต้องหาภาพประกอบ ใส่ข้อความบรรยายต่าง ๆ และในการนำเสนอแต่ละครั้งมีทั้งเรื่องที่เป็นเรื่องทั่วไปหรือเรื่องเฉพาะกลุ่ม ด้วยความที่ไม่รู้ว่าผู้ฟังจะเข้าใจในแบบที่เราเข้าใจไหมไม่รู้ว่าจะเอาภาพแบบไหนมาอธิบาย หรือผู้ฟังมีคำถามที่เราไม่ได้เตรียมอะไรไว้อธิบาย ซึ่งนอกจากต้องใช้เวลาในการเตรียมการนำเสนอ การจัดลำดับเรื่อง และเรื่องที่อาจจะไม่ได้เตรียมไว้

คงจะดีกว่าถ้าเราสามารถอธิบายสิ่งที่อยู่ในหัวเราให้ออกมาเป็นภาพได้เลย ไม่ต้องเตรียมอะไรเป็นพิเศษ ให้ผู้ฟังเห็นภาพพร้อมผู้พูด เพราะสิ่งสำคัญของการสื่อสารคือการเห็นภาพร่วมกัน เพื่อความเข้าใจที่ตรงกัน

ดังนั้นเทคโนโลยีนี้จะเข้าไปช่วยสำหรับผู้พูดบรรยายต่าง ๆ ได้ทั้งสิ้น ในห้องเรียนที่คุณครูพูดสอนแล้วเด็กเห็นภาพ และเด็กได้เห็นภาพที่ตัวเองพูดออกมาซึ่งจะเพิ่มบรรยากาศในห้องเรียน การพูดนำเสนอในที่สาธารณะ ถึงคนที่อยู่ไกลไม่ค่อยได้ยินเสียงก็สามารถเห็นภาพแล้วอาจจะอยากเข้ามาฟัง การเล่านิทาน หรือแม้แต่เรื่องต่าง ๆ ที่แค่อยากเล่าออกไป ก็จะมีภาพประกอบเพิ่มความรู้สึกไปกับเรื่องที่เล่า

โดยผู้ใช้ระบบ จะพูดในเรื่องใดเรื่องหนึ่ง เมื่อระบบฟังเสียงได้จะฟังเสียงและแสดงภาพของคำที่ได้ยิน ผู้ใช้สามารถสั่งการด้วยเสียงเพื่อให้ภาพเป็นไปตามที่ต้องการ เช่น การขยับภาพ ลบภาพทั้งหมดที่แสดงออกมา ผู้ใช้สามารถเพิ่มระดับการแสดงภาพโดยเพิ่มระดับความเข้าใจความหมาย ระบบสามารถแสดงภาพตามความหมายโดยตรงหรือโดยอ้อมได้ ภาพที่แสดงออกมาจะเป็นแบบในอุดมคติที่สุด ผู้ใช้สามารถเพิ่มภาพสื่อต่าง ๆ ที่ตัวเองต้องการได้สามารถเดิ่มจำนวนภาพที่ใส่เข้าไปได้ โดยการเพิ่มจะเป็นการใส่ภาพและคำที่เกี่ยวข้องให้จดจำภาพที่ใส่แทนภาพนั้น ผู้ใช้สามารถใส่ไฟล์เสียงแล้วให้ระบบเล่นเป็นวิดีโอ ผู้ใช้สามารถบันทึกวิดีโอระหว่างพูดได้

**คุณสมบัติสำคัญของระบบ**

1. สามารถฟังเสียงภาษาไทยได้
2. ระบบสามารถรับทั้งไฟล์เสียงและแบบ Real-time ได้
3. ระบบสามารถแสดงภาพตามที่ระบบเข้าใจ
4. สามารถกำหนดภาพที่ต้องการแสดงโดยเฉพาะเองได้
5. สามารถอัดบันทึกวิดีโอได้

**ประโยชน์ที่จะได้รับ**

1. การพูดบรรยายต่าง ๆ จะไม่ต้องใช้เวลานานในการเตรียมนำเสนอ
2. ทุกคนจะเห็นภาพความคิดของกันและกัน ด้วยความเข้าใจที่ตรงกัน
3. ใช้ภาพแทนการสื่อสารกับคนพิการทางการได้ยิน

**เงื่อนไขยองระบบ**

1. ระบบต้องได้รับเสียงที่ชัดเจน
2. เสียงที่รับมาจะต้องตรงกับคำที่ระบบเรียนรู้
3. ระบบต้องเรียนรู้ จากการรับภาพกับเสียง
4. ระบบต้องเข้าใจภาษา ว่าเหตุการณ์หนึ่งสามารถพูดได้หลายประโยค

**เทคโนโลยีที่คาดว่าจะใช้ หรือเกี่ยวข้อง**

1. NLP
2. Speech-to-text
3. Text-to-image
4. Api คำ ภาษาไทย
5. Api ภาพ
6. Code

**เอกสารอ้างอิง**

* [**https://www.sify.com/ai-analytics/what-are-text-to-video-models-in-ai-and-how-they-are-changing-the-world**](https://www.sify.com/ai-analytics/what-are-text-to-video-models-in-ai-and-how-they-are-changing-the-world)
* [**https://alltechnews.medium.com/insights-into-text-to-video-ai-technology-how-it-works-83507b5b2081**](https://alltechnews.medium.com/insights-into-text-to-video-ai-technology-how-it-works-83507b5b2081)
* [**https://www.appypie.com/blog/ai-text-to-video-generation**](https://www.appypie.com/blog/ai-text-to-video-generation)