

# Multimodal AI Studio Platform

*Nền tảng phân tích và sáng tạo nội dung đa thể thức*

Trung tâm Dịch vụ Truyền hình

Trung tâm Phân tích Dữ liệu



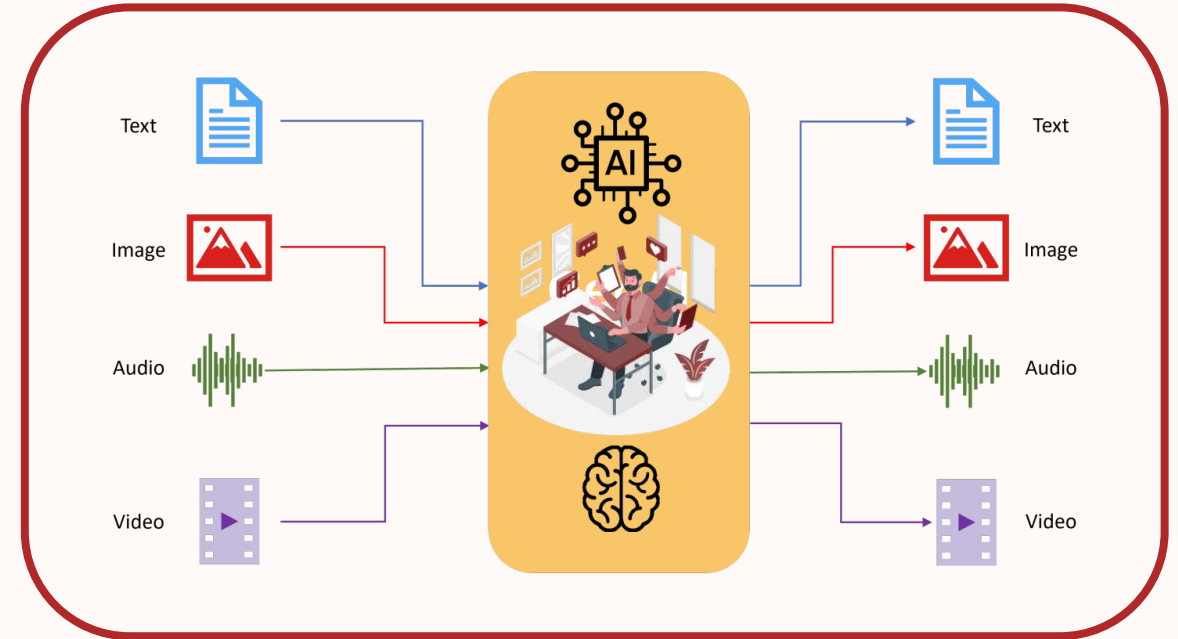
# Mục lục

- 1 Nội dung ý tưởng
- 2 Phân tích thị trường
- 3 Đề xuất tính năng, giao diện
- 4 Kịch bản triển khai

# Nền tảng phân tích, sáng tạo nội dung đa thể thức

Xây dựng Platform hỗ trợ phân tích và tạo nội dung đa thể thức:

- Tích hợp sẵn các tính năng hỗ trợ phân tích, sáng tạo nội dung theo yêu cầu, mô tả.
- Có giao diện web/app thân thiện cho người sử dụng, dễ dàng thao tác
- Phát huy tối đa tính sáng tạo của con người.
- Hỗ trợ đa ngôn ngữ: tiếng Anh, tiếng Việt, ...



## ★ Big Idea

“Đơn giản, dễ tiếp cận, hiệu quả nhanh”

## Key Message

“Hãy đưa chúng tôi ý tưởng của bạn, chúng tôi sẽ giúp bạn hiện thực hoá”

## Tính năng nổi bật

- Tạo văn bản, hình ảnh, audio, video chất lượng
- Hỗ trợ tạo tài liệu trình chiếu, video đào tạo
- Biên tập, tạo video ngắn, highlight, best cut từ nội dung video gốc
- Phân tích, tổng hợp, tóm tắt, truy vấn, số hóa nội dung văn bản
- Hỗ trợ ý tưởng, sáng tác nhạc
- ...



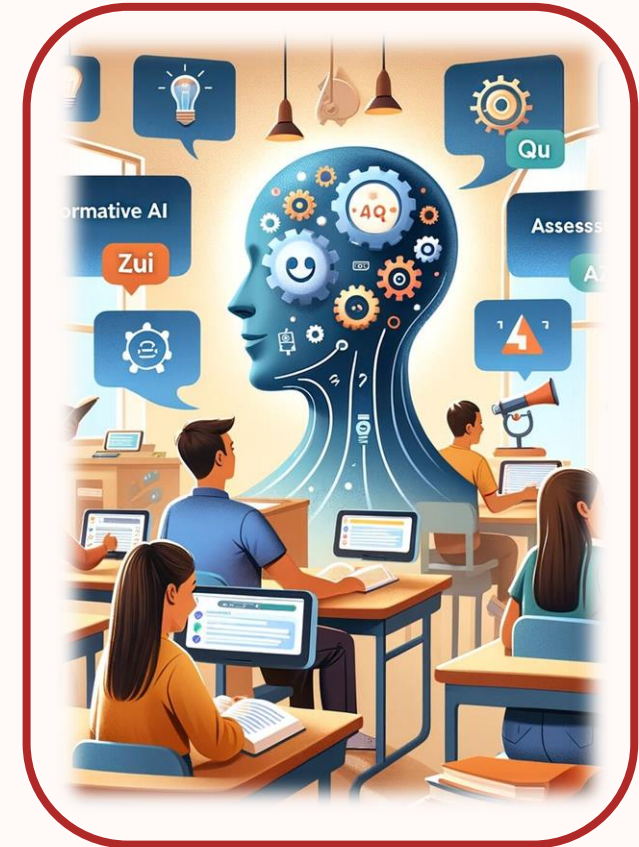
# Người dùng mục tiêu

## 1 Nội bộ:

Truyền thông, văn phòng, dịch vụ truyền hình, dịch vụ khách hàng, nhân viên bán hàng, ...

## 2 Các đối tượng cá nhân tổ chức:

- Hỗ trợ tạo tài liệu trình chiếu, video đào tạo
- Người sáng tạo nội dung: youtuber, tik token, Facebook user, Podcaster, ...
- Người review phim, audio book, video...
- Thư ký văn phòng: Viết tổng kết các cuộc họp, ...
- Nhân viên văn phòng
- Nhân sự truyền thông: viết bài, làm video, ...
- Nhân sự chăm sóc khách hàng
- Nhân viên điều hành bán hàng, người bán hàng online: làm video giới thiệu sản phẩm, livestream bán hàng,
- Giảng viên, học sinh, sinh viên: Làm video bài giảng, tóm tắt nội dung bài học, ...



# Một số Use Case

1

## Tạo nội dung

- Video truyền thông, bán hàng, chăm sóc khách hàng, cá nhân hoá theo người dùng
- Sinh ảnh, tạo bài viết, audio, video, tạo nhạc từ văn bản
- Sáng tạo nội dung theo mô tả: bài giảng, slide
- Tạo video phục vụ livestream bán hàng

2

## Phân tích và chuyển đổi

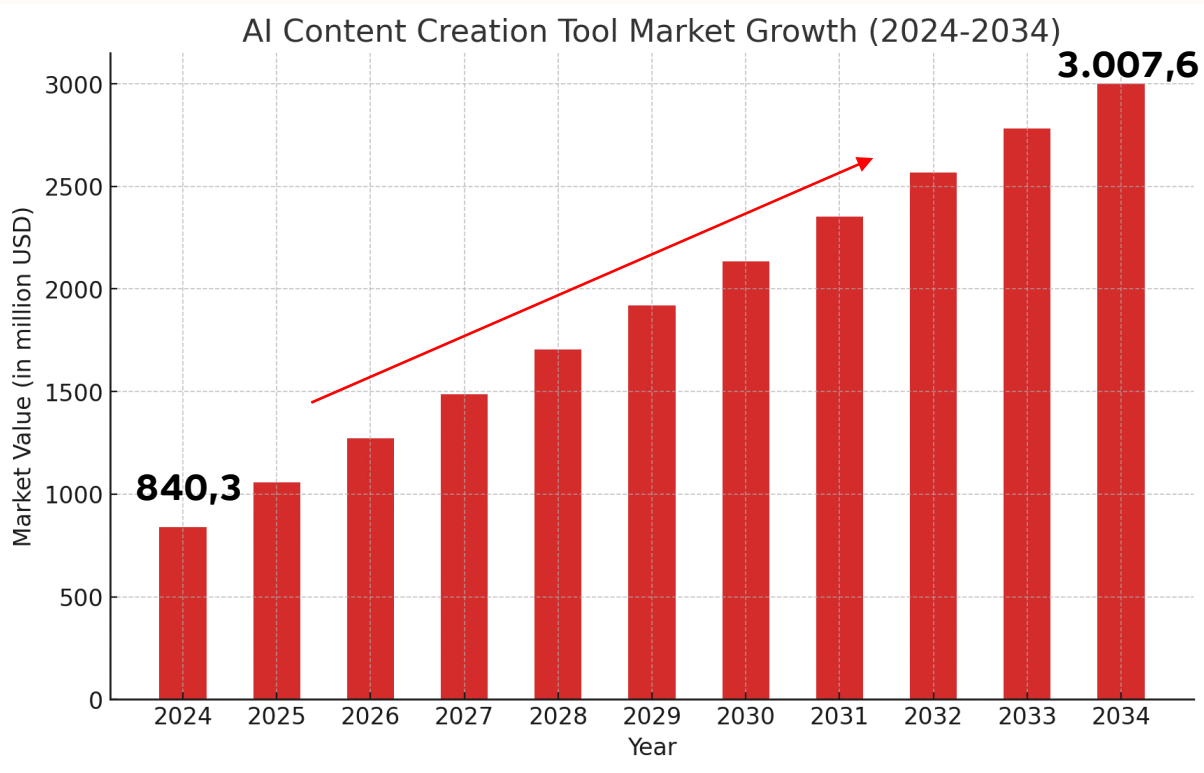
- Phân tích audio, video: nhận diện cảm xúc, tóm tắt, trích xuất thông tin quan trọng
- Chuyển giọng nói thành văn bản, văn bản thành giọng nói
- Video Translation: dịch nội dung, chuyển đổi giọng nói kết hợp khẩu hình miệng
- Tóm tắt, highlight video, phim
- Trích xuất, phân tích văn bản: OCR, tóm tắt, hỏi đáp, phân tích báo cáo tài chính

3

## Chỉnh sửa và tùy chỉnh

- Tạo video custom: hình ảnh, giọng nói, nội dung cá nhân hóa
- Hỗ trợ tạo và chỉnh sửa nội dung, tăng chất lượng podcast
- Thêm, sửa, xóa đối tượng trong hình ảnh/video, tăng cường chất lượng

# Nghiên cứu thị trường



Thị trường công cụ tạo nội dung bằng AI dự kiến đạt giá trị **840,3 triệu USD** vào năm 2024 và sẽ tăng trưởng với tốc độ **CAGR 13,60%** từ 2024 đến 2034, đạt **3.007,6 triệu USD** vào năm **2034**.

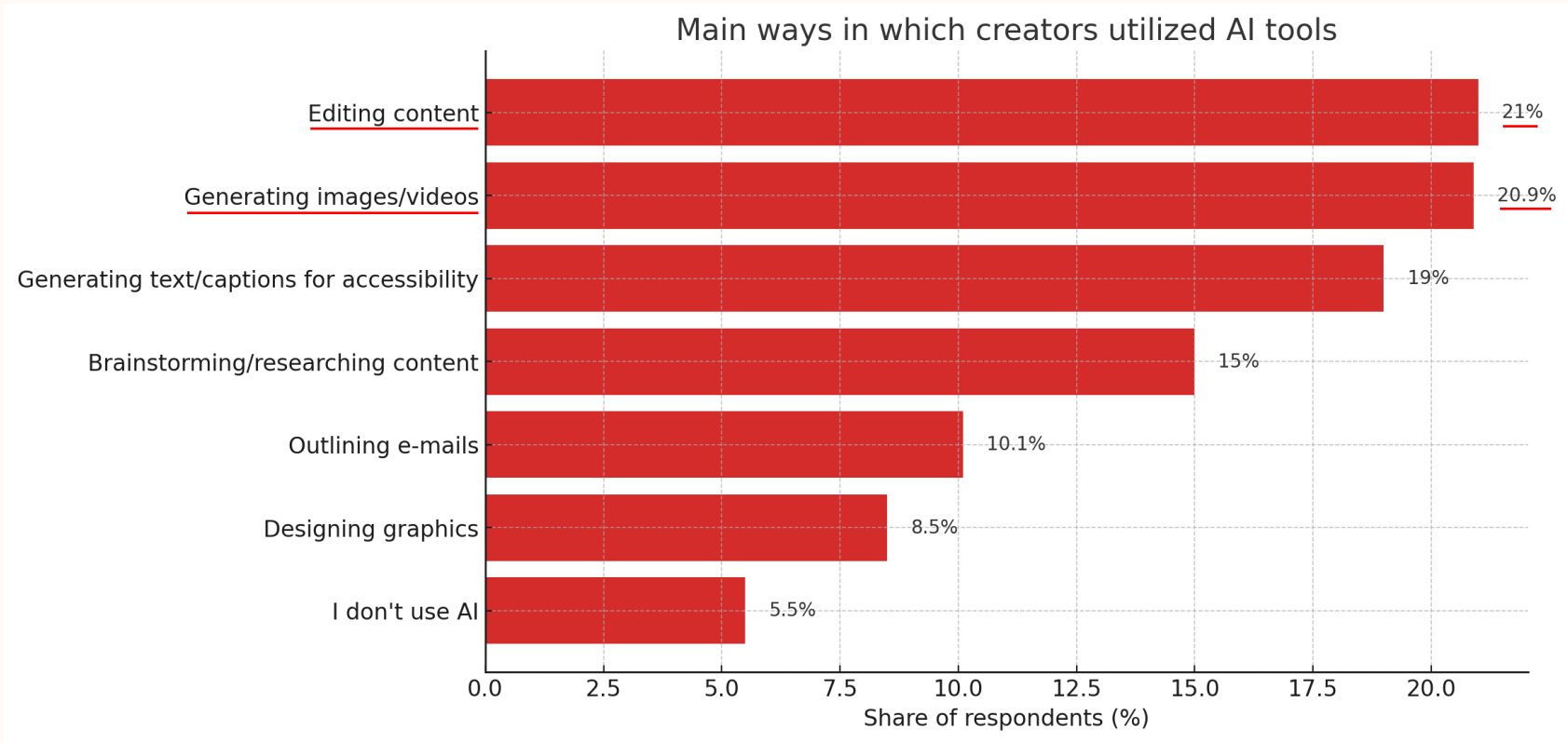
&

Theo idc-a.org, đến năm **2026, 90%** nội dung online được tạo ra có sự hỗ trợ của AI

*Theo Future Market Insights*

1

# Nghiên cứu thị trường



Theo Statista

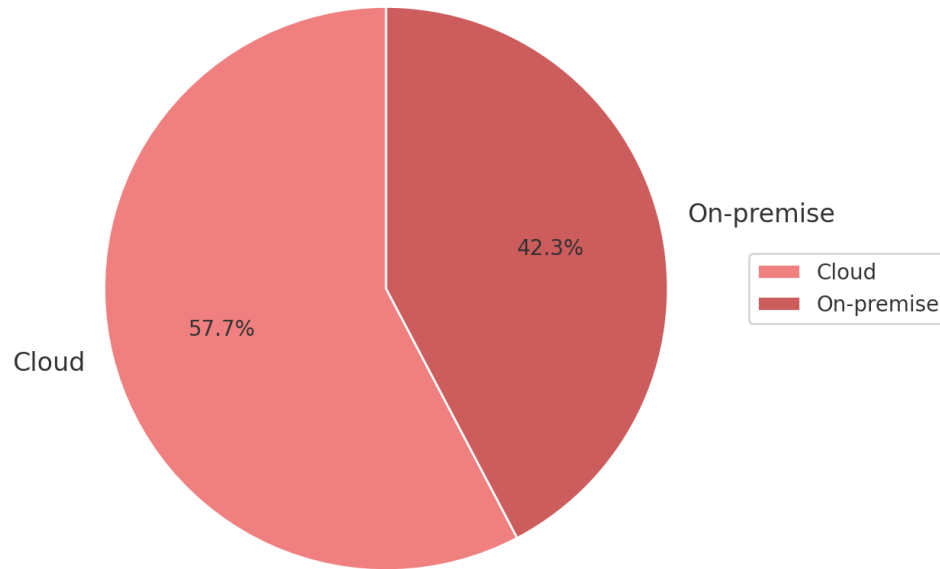
→ Có **nhu cầu mạnh mẽ** đối với các công cụ có thể tăng năng suất trong việc **chỉnh sửa** và **tạo nội dung** (hình ảnh và văn bản), cũng như hỗ trợ khả năng tiếp cận và phát triển ý tưởng.

2



# Nghiên cứu thị trường

Global AI Powered Content Creation Market 2024-2033 (By Deployment Mode)



Cho thấy sự **ưu tiên** cho các **giải pháp Cloud**.



Do tính linh hoạt, khả năng mở rộng và khả năng truy cập mà các dịch vụ cloud mang lại, đặc biệt có lợi cho các doanh nghiệp với môi trường làm việc từ xa hoặc hybrid.

*Theo Customer Market Insights*

3

# Đối thủ cạnh tranh



## Youtube Short

Cho phép người dùng đăng video ngắn, nhưng không có công cụ sáng tạo nội dung



## Tiktok

Nền tảng sáng tạo nội dung lớn, nhưng vấn đề bản quyền chưa được bảo đảm



## AI Tools

Các công cụ được các nhà cung cấp đơn lẻ theo 1 nhóm chức năng nhất định. Chưa có công cụ kết hợp các nhóm tính năng



## TV360

Đáp ứng bản quyền nội dung về thể thao, phim, truyền hình. Có thể tạo mới Tab TV360 Short Video với tính năng sáng tạo nội dung và chỉnh sửa đa thể thức

# Phân tích người dùng mục tiêu

## 1 Nội bộ Viettel:

- **Quảng cáo, truyền thông phục vụ bán hàng, chăm sóc khách hàng:** Tỷ lệ chuyển đổi có thể tăng từ 3-5% lên đến 13-15%.
- **Đào tạo, truyền thông nội bộ:** Dưới dạng văn bản, giọng nói, video, hình ảnh.

## 2 Người sáng tạo nội dung:

Việt Nam hiện đang có **200.000 Content Creator** hoạt động trên mạng xã hội, với hàng triệu content được sản xuất mỗi ngày. → Có một nhu cầu lớn về việc tạo nội dung nhanh chóng và hiệu quả để chia sẻ trên mạng xã hội.

## 3 Doanh nghiệp:

Có nhu cầu **sáng tạo nội dung** để thu hút khách hàng như: tạo nội dung truyền thông và đào tạo.



# Pain points của người dùng

- Nhu cầu tạo nội dung số, đặc biệt là nội dung đa phương tiện, ngày càng tăng.
- Tại Việt Nam chưa có nhiều các nền tảng, hoặc các nền tảng vẫn đang cung cấp chỉ 1 hoặc 1 vài tính năng riêng biệt với nhau.

## Kết luận

- ➔ Xây dựng 1 platform cho phép người dùng phân tích, sáng tạo nội dung đa thể thức dựa trên giải pháp đám mây (cloud).
- ➔ Tích hợp nhiều tiện ích để rút ngắn thời gian phân tích, sản xuất nội dung.



# Tổng quan sản phẩm

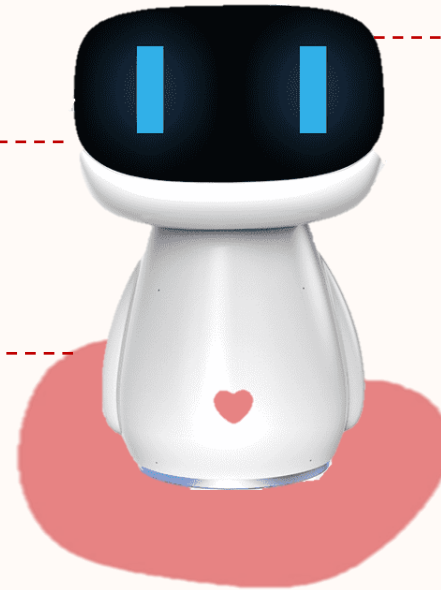
## Tôn chỉ

ĐƠN GIẢN – LINH HOẠT – HIỆU QUẢ

## USP

Hỗ trợ người dùng tạo, chỉnh sửa nội dung đa thể thức một cách **nhẹ nhàng và dễ dàng.**

**Đảm bảo an toàn** cho nội dung của người sáng tạo, đồng thời cung cấp thư viện nội dung đa dạng, chất lượng cao.



## Target users

- Nội bộ Viettel
- Đối tác của Viettel
- Cá nhân sáng tạo nội dung
- Doanh nghiệp

## Lợi thế cạnh tranh

Tập trung vào thị trường Việt Nam, hiểu rõ tâm lý và nhu cầu của người dùng.

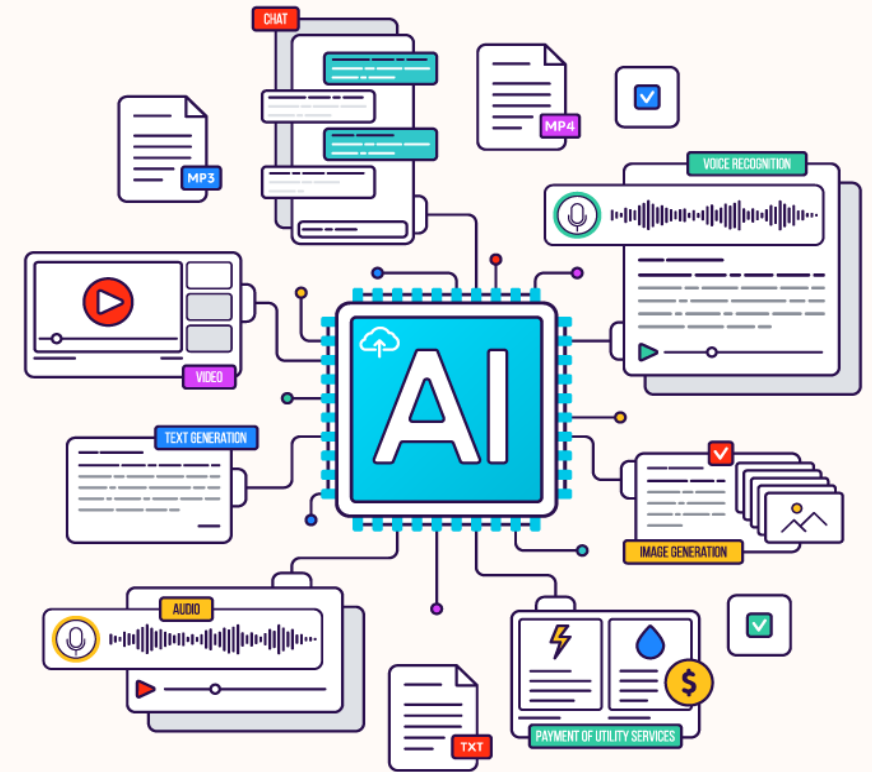
Kết hợp nhiều tính năng hữu ích trong một nền tảng duy nhất.

Giao diện thân thiện, dễ dàng thao tác, phù hợp với cả người mới bắt đầu.

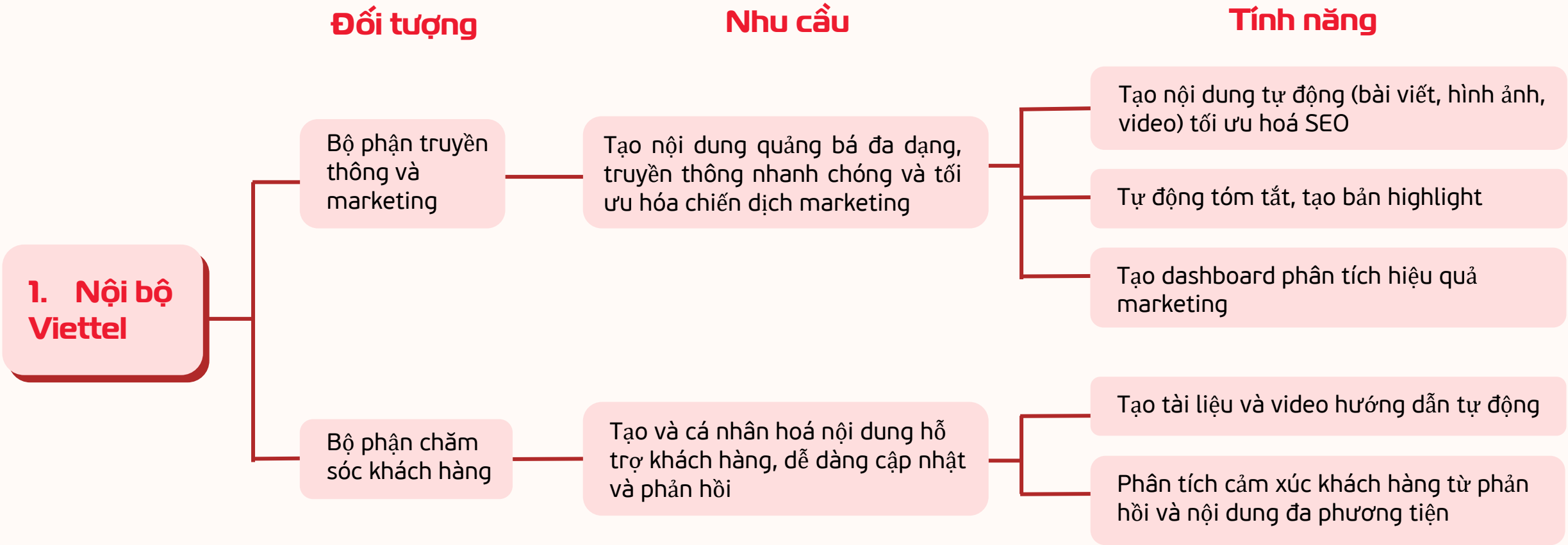
# Nền tảng công nghệ

## Platform sẽ được tích hợp các công nghệ AI:

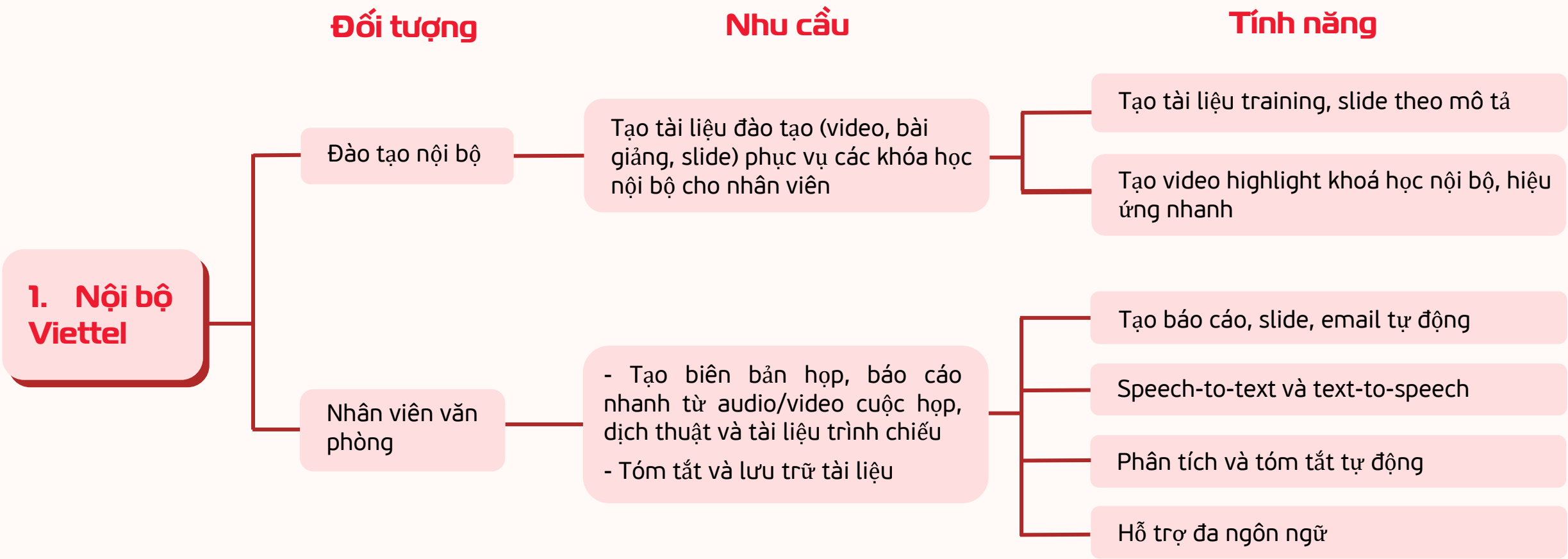
- Mô hình ngôn ngữ lớn (Large Language Model)
- Multimodal Large Language Model
- Generative AI: hình ảnh, âm thanh
- Xử lý ảnh, video, xử lý ngôn ngữ tự nhiên
- MC ảo
- Có thể tích hợp mở rộng với các nền tảng như Chat GPT, Gemini, Copilot, Mid journey...



# Đề xuất tính năng theo phân khúc

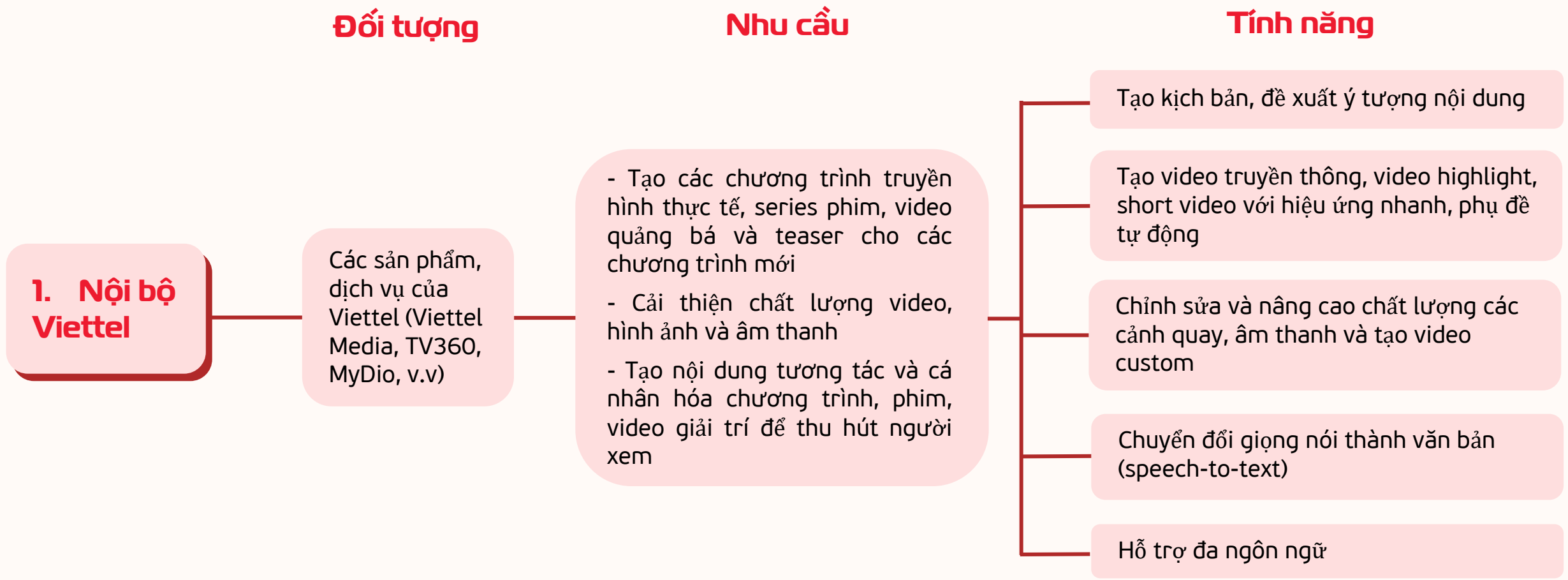


# Đề xuất tính năng theo phân khúc

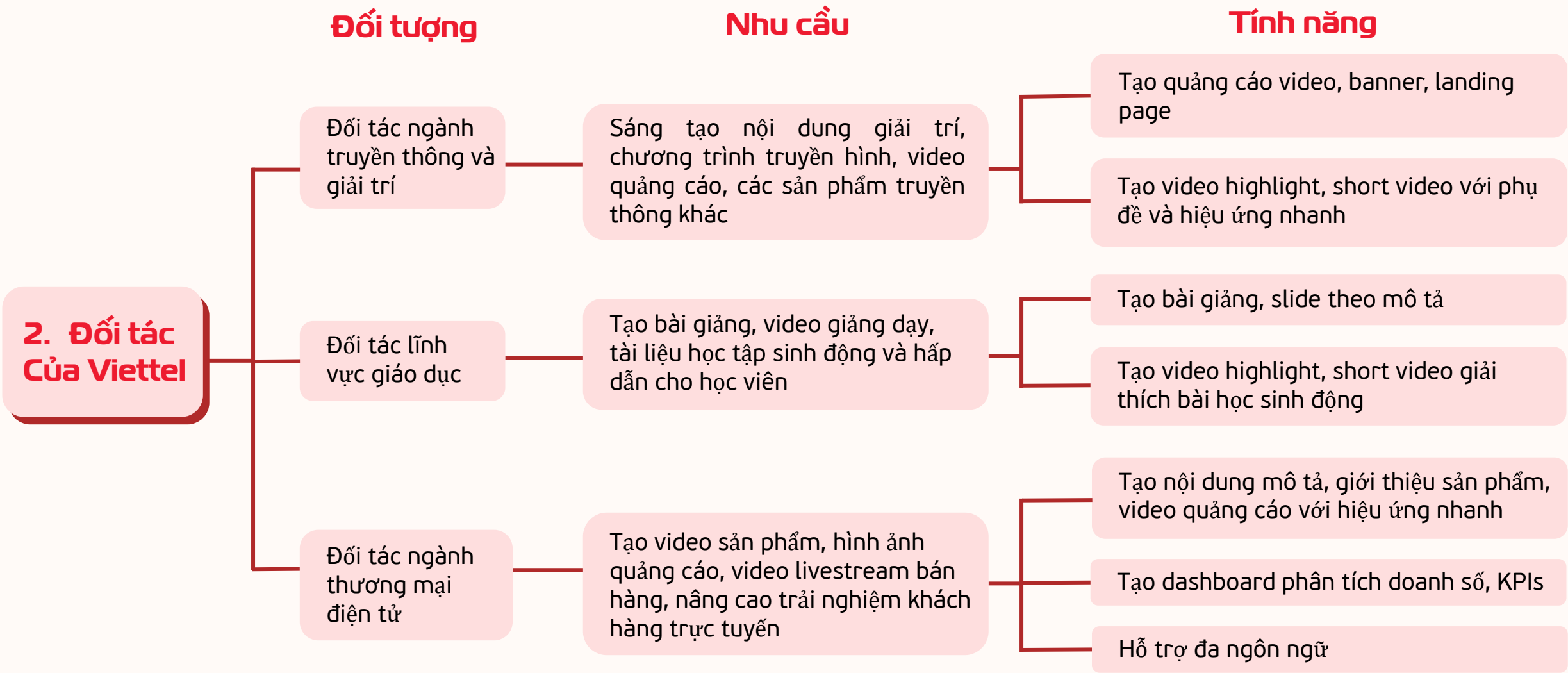




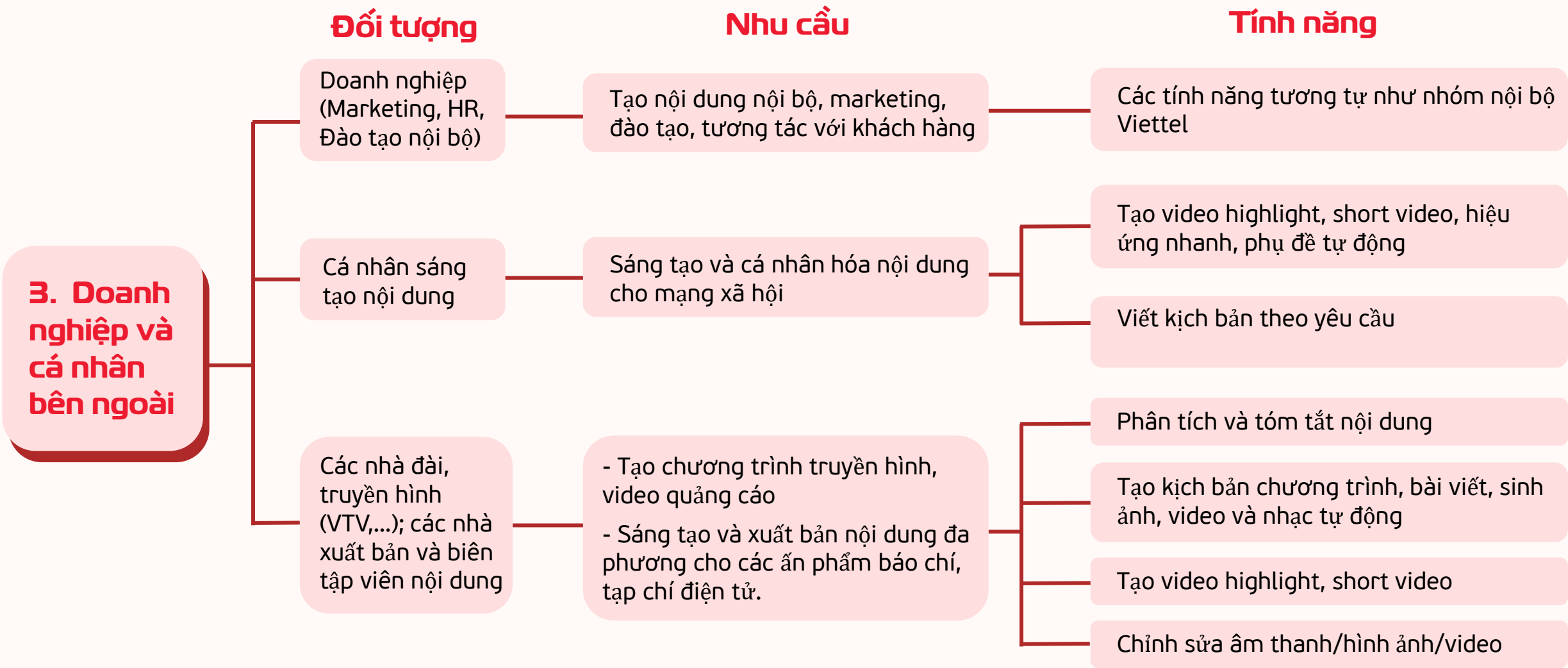
# Đề xuất tính năng theo phân khúc



# Đề xuất tính năng theo phân khúc



# Đề xuất tính năng theo phân khúc



# Phương thức tương tác và giao diện

Nhóm tính năng	Tính năng	Kênh triển khai	Giao diện	Phương thức tương tác		Ví dụ Input	Ví dụ Output
				Input	Output		
Tạo nội dung	Tạo video truyền thông, quảng bá sản phẩm, CSKH	Web, App Mobile, API	<ul style="list-style-type: none"><li>- Dạng drag-and-drop hình ảnh, video và chọn hiệu ứng từ thư viện.</li><li>- Hỗ trợ preview trực tiếp để xem thay đổi ngay lập tức.</li><li>- Các thanh công cụ tối giản để chỉnh sửa hiệu ứng, chọn mẫu template, thêm phụ đề với AI gợi ý v.v</li></ul>	Tạo nội dung bằng kéo-thả dựa trên văn bản mô tả, hình ảnh, video clip, chọn hiệu ứng, thêm template và phụ đề	Video hoàn chỉnh với hiệu ứng, nhạc nền, phụ đề	<ul style="list-style-type: none"><li>- Mô tả: 'Video quảng bá sản phẩm A'</li><li>- Kéo – thả hình ảnh: 'Hình ảnh sản phẩm'</li><li>- Chọn hiệu ứng 'fade-in'</li><li>- Mẫu template: số 3</li><li>- Thêm phụ đề: 'Có'</li></ul>	Video quảng bá sản phẩm với hiệu ứng chuyển động và nhạc nền
	Tạo bài viết, kịch bản, sinh ảnh, nhạc và slide theo mô tả dạng text (command)	Web, App Mobile, API	<ul style="list-style-type: none"><li>- Trường nhập văn bản với gợi ý prompt thông minh (AI sẽ hỗ trợ đề xuất).</li><li>- Hỗ trợ tùy chọn chọn style (phong cách) hoặc theme (chủ đề) từ menu dropdown.</li></ul>	Nhập văn bản hoặc chọn template	Nội dung tự động được sinh từ văn bản, bao gồm bài viết, kịch bản, ảnh, slide hoặc nhạc	Nhập văn bản: 'Bài viết về AI trong marketing'	Bài viết hoàn chỉnh về AI, kèm theo hình ảnh minh họa

# Phương thức tương tác và giao diện

Nhóm tính năng	Tính năng	Kênh triển khai	Giao diện	Phương thức tương tác		Ví dụ Input	Ví dụ Output
				Input	Output		
Phân Tích và Chuyển Đổi	Phân tích nội dung, cảm xúc, hành vi từ audio/video	Web, App Mobile, API	Dashboard phân tích nội dung, cảm xúc, hành vi từ audio/video, hiển thị kết quả qua biểu đồ và báo cáo chi tiết	Tải lên audio hoặc video	Báo cáo nội dung, cảm xúc, hành vi với biểu đồ phân tích	Tải lên file audio từ cuộc gọi khách hàng	Báo cáo cảm xúc, nội dung yêu cầu, hành vi của khách hàng trong cuộc gọi
	Chuyển giọng nói thành văn bản (speech-to-text)	Web, App Mobile, API	<ul style="list-style-type: none"><li>- Trường nhập ghi âm hoặc khu vực tải tệp (kéo thả)</li><li>- Hiển thị văn bản chuyển đổi với tính năng highlight và chỉnh sửa trực tiếp</li></ul>	Ghi âm giọng nói trực tiếp hoặc tải file audio, video	Văn bản chuyển đổi từ giọng nói, có thể chỉnh sửa	Ghi âm giọng nói cuộc họp	Văn bản được chuyển đổi từ giọng nói trong cuộc họp
	Chuyển văn bản thành giọng nói (text-to-speech)	Web, App Mobile, API	<ul style="list-style-type: none"><li>- Chọn giọng nói mô phỏng với các tùy chỉnh như: ngữ điệu, giọng nam/nữ, tốc độ đọc</li><li>- Kết quả phát trực tiếp ngay trong giao diện với nút replay hoặc tải về định dạng file audio</li></ul>	Tải lên văn bản	Giọng nói được chuyển đổi từ văn bản	<ul style="list-style-type: none"><li>- Nhập văn bản: 'Chào mừng đến với video hướng dẫn'</li><li>- Tùy chỉnh giọng: 'Nữ'</li><li>- Tốc độ: 'Trung bình'</li></ul>	Giọng nói mô phỏng: 'Chào mừng đến với video hướng dẫn'

# Phương thức tương tác và giao diện

Nhóm tính năng	Tính năng	Kênh triển khai	Giao diện	Phương thức tương tác		Ví dụ Input	Ví dụ Output
				Input	Output		
Phân Tích và Chuyển Đổi	Dịch nội dung video, chuyển đổi giọng nói kết hợp khẩu hình miệng	Web, App Mobile, API	<ul style="list-style-type: none"><li>Hiển thị khung video preview, nơi người dùng nhập ngôn ngữ đầu vào và đầu ra</li><li>Tính năng chọn style giọng nói, tốc độ dịch và điều chỉnh thời gian khớp khẩu hình miệng</li></ul>	Tải lên video gốc, ngôn ngữ nguồn	Video đã dịch với giọng nói và khẩu hình miệng đồng bộ	Tải lên video tiếng Anh, muốn chuyển sang tiếng Việt	Video đã dịch sang tiếng Việt, giọng nói và khẩu hình miệng đồng bộ
	Tóm tắt, highlight video, phim	Web, App Mobile, API	<ul style="list-style-type: none"><li>Hiển thị timeline video với các đoạn tóm tắt hoặc highlight được tự động đề xuất</li><li>Tích hợp tính năng chọn các đoạn cụ thể để chỉnh sửa hoặc xóa bỏ nếu không cần thiết</li></ul>	Tải lên video gốc	Video tóm tắt, highlight hoặc phân đoạn đặc biệt	Tải lên video phim tài liệu	Video đã được tóm tắt với những phân đoạn quan trọng
	Phân tích, tạo dashboard từ số liệu và tóm tắt văn bản	Web, App Mobile, API	<ul style="list-style-type: none"><li>Khu vực kéo-thả trực quan với nút tải file (PDF, hình ảnh, văn bản, excel)</li><li>Nút tóm tắt (Quick Summary), hỏi đáp (Q&amp;A), phân tích (Analyze) và tạo Dashboard</li></ul>	Tải lên hình ảnh tài liệu, PDF, văn bản, số liệu	Kết quả phân tích, tóm tắt, hoặc trích xuất thông tin quan trọng	Tải lên báo cáo tài chính	Tóm tắt thông tin quan trọng từ báo cáo tài chính

# Phương thức tương tác và giao diện

Nhóm tính năng	Tính năng	Kênh triển khai	Giao diện	Phương thức tương tác		Ví dụ Input	Ví dụ Output
				Input	Output		
Chỉnh Sửa và Tùy Chỉnh	Chỉnh sửa âm nhạc, giọng nói, voice-over và video	Web, App Mobile, API	<ul style="list-style-type: none"><li>Dạng editor chuyên dụng có thể chỉnh sửa từng lớp (layer) như âm thanh, hình ảnh, phụ đề</li><li>Thư viện tích hợp với các hiệu ứng âm thanh (ví dụ: robotic, reverb) và các tùy chọn tùy chỉnh sâu</li></ul>	Tải lên file audio	File audio đã chỉnh sửa với chất lượng âm thanh tốt hơn	<ul style="list-style-type: none"><li>Tải lên file audio gốc</li><li>Hiệu ứng âm thanh: 'Robotic'</li></ul>	File audio đã được thêm hiệu ứng
	Tạo video custom, hình ảnh, giọng nói, nội dung cá nhân hóa	Web, App Mobile, API	<ul style="list-style-type: none"><li>Giao diện editor cho phép tùy chỉnh sâu từng yếu tố trong video: thay đổi màu sắc, hiệu ứng, chèn hình ảnh hoặc văn bản</li><li>Tính năng preview và xuất bản ngay trong giao diện</li></ul>	Tải lên video, hình ảnh, giọng nói, văn bản	Video hoặc hình ảnh chỉnh sửa, nội dung cá nhân hóa	<ul style="list-style-type: none"><li>Tải lên video quảng cáo sản phẩm A</li><li>Tùy chỉnh giọng: 'Thêm thuyết minh'</li><li>Tùy chỉnh đối tượng: 'Xoay ngang'</li></ul>	Video quảng cáo với thuyết minh và sự cá nhân hoá
	Thêm, sửa, xóa đối tượng trong hình ảnh/video	Web, App Mobile, API	<ul style="list-style-type: none"><li>Công cụ chỉnh sửa đối tượng với các tùy chọn tự động nhận diện và xóa bỏ đối tượng cụ thể.</li><li>Cho phép khoanh vùng trực tiếp trên hình ảnh/video để chỉnh sửa</li></ul>	Hình ảnh, video, đối tượng	Hình ảnh hoặc video chỉnh sửa với đối tượng thay đổi	<ul style="list-style-type: none"><li>Tải lên hình ảnh sản phẩm</li><li>Tùy chỉnh đối tượng: 'Xóa hình người'</li></ul>	Hình ảnh sản phẩm với đối tượng đã được xóa bớt

# Mô hình kinh doanh

- 1 **Cung cấp gói dịch vụ theo nhu cầu của khách hàng:**
  - Theo từng nhóm tính năng (voice, hình ảnh, video, text,...)
  - Theo toàn bộ tính năng
  - Thu phí theo tháng/năm/trọn đời
- 2 **Bundle, bán kèm các dịch vụ SME/dịch vụ đại trà** (như Hoá đơn điện tử, vContract, My Sign, TV360, My Box, Cloud...).
- 3 Có thể cung cấp sẵn **gói giới hạn miễn phí cho các KH đang sử dụng dịch vụ của Viettel** để giữ chân khách hàng, nếu sử dụng hết giới hạn miễn phí thì cần nạp thêm tiền vào tài khoản.
- 4 Với KH chưa sử dụng dịch vụ nào của Viettel có thể được **ưu đãi** nếu sử dụng kèm các dịch vụ CNTT, viễn thông.
- 5 **Chiết khấu cho hợp đồng dài hạn**





# Cấu trúc chung các gói dịch vụ

## 4 gói chính

- Gói Generator:** Tạo nội dung (video, văn bản, phụ đề, highlight, v.v. theo mô tả).
- Gói Transformer:** Phân tích và chuyển đổi (speech-to-text, text-to-speech, nhận diện cảm xúc, tóm tắt, highlight nội dung, v.v.).
- Gói Editor:** Chỉnh sửa và tùy chỉnh (video, podcast, hình ảnh, nội dung đa phương tiện).
- Gói Ultimate:** Bao gồm tất cả tính năng từ 3 gói trên.

# Cấu trúc chung các gói dịch vụ

## 1 Phân chia theo thời gian:

- **Theo tháng:** Giá tiêu chuẩn.
- **Theo năm:** Giảm 20% (khuyến khích đăng ký lâu dài).

## 2 Cấp độ trong từng gói:

- **Basic:** Dành cho người dùng cá nhân hoặc doanh nghiệp nhỏ.
- **Standard:** Dành cho doanh nghiệp vừa với nhu cầu thường xuyên.
- **Premium:** Dành cho doanh nghiệp lớn hoặc tổ chức cần sử dụng toàn diện và không giới hạn.



# Gói Generator

Tính năng:

- Tạo video truyền thông, quảng bá sản phẩm, CSKH
- Tạo bài viết, sinh ảnh, nhạc và slide theo mô tả dạng text (command)

Cấp độ	Theo tháng (VNĐ)	Theo năm (VNĐ)	Hạn mức sử dụng
Basic	3,000,000	28,800,000	50 phút AI tạo nội dung/tháng, 80 templates, 100GB lưu trữ, không giới hạn xuất bản, 2 giọng nói, tối đa 1 người dùng
Standard	6,000,000	57,600,000	150 phút AI tạo nội dung/tháng, 200 templates, 250GB lưu trữ, không giới hạn xuất bản, 5 giọng nói, tối đa 5 người dùng
Premium	10,000,000	96,000,000	Không giới hạn AI tạo nội dung, 500 templates, 1TB lưu trữ, không giới hạn xuất bản, 10 giọng nói, tối đa 20 người dùng

# Gói Transformer

Tính năng:

- Phân tích nội dung, cảm xúc, hành vi từ audio/video
- Chuyển giọng nói thành văn bản (speech-to-text)
- Chuyển văn bản thành giọng nói (text-to-speech)
- Dịch nội dung video, chuyển đổi giọng nói kết hợp khẩu hình miệng
- Tóm tắt, highlight video, phim
- Trích xuất và phân tích văn bản (OCR, tóm tắt, hỏi đáp, phân tích báo cáo tài chính)

Cấp độ	Theo tháng (VNĐ)	Theo năm (VNĐ)	Hạn mức sử dụng
Basic	3,000,000	28,800,000	50 phút phân tích audio/video/tháng, tối đa 100 tệp văn bản, 100GB lưu trữ, 1 người dùng
Standard	6,000,000	57,600,000	150 phút phân tích audio/video/tháng, tối đa 300 tệp văn bản, 250GB lưu trữ, tối đa 5 người dùng
Premium	10,000,000	96,000,000	Không giới hạn phân tích audio/video và tệp văn bản, 1TB lưu trữ, hỗ trợ báo cáo chi tiết, tối đa 20 người dùng

# Gói Editor

Tính năng:

- Chỉnh sửa âm nhạc, giọng nói, voice-over và video
- Tạo video custom, hình ảnh, giọng nói, nội dung cá nhân hóa
- Thêm, sửa, xóa đối tượng trong hình ảnh/video

Cấp độ	Theo tháng (VNĐ)	Theo năm (VNĐ)	Hạn mức sử dụng
Basic	3,000,000	28,800,000	50 phút chỉnh sửa video/audio, 10 video hoặc 15 audio/tháng, 100GB lưu trữ, không giới hạn xuất bản
Standard	6,000,000	57,600,000	150 phút chỉnh sửa video/audio, 30 video hoặc 35 audio/tháng, 250GB lưu trữ, tối đa 5 người dùng
Premium	10,000,000	96,000,000	Không giới hạn chỉnh sửa video/audio, 1TB lưu trữ, hỗ trợ chỉnh sửa nâng cao

# Gói Ultimate

Tính năng:

- Bao gồm tất cả tính năng từ 3 gói trên

Cấp độ	Theo tháng (VNĐ)	Theo năm (VNĐ)	Hạn mức sử dụng
Basic	5,000,000	48,000,000	50 phút AI tạo nội dung/tháng, 80 templates, 50 phút phân tích/chỉnh sửa audio/video/tháng, tối đa 100 tệp văn bản, 100GB lưu trữ, không giới hạn xuất bản, 2 giọng nói, tối đa 1 người dùng
Standard	10,000,000	96,000,000	150 phút AI tạo nội dung/tháng, 200 templates, 150 phút phân tích/chỉnh sửa audio/video/tháng, tối đa 300 tệp văn bản, 250GB lưu trữ, không giới hạn xuất bản, 5 giọng nói, tối đa 5 người dùng
Premium	18,000,000	172,800,000	Không giới hạn AI tạo nội dung, 500 templates, không giới hạn phân tích/chỉnh sửa audio/video và tệp văn bản, 1TB lưu trữ, không giới hạn xuất bản, 10 giọng nói, tối đa 20 người dùng

# Ưu đãi kèm theo

- ① **Giảm 20%** cho tất cả các gói **đăng ký theo năm**
- ② **Dùng thử miễn phí:**
  - **Basic:** 7 ngày
  - **Standard:** 15 ngày (cho khách hàng lớn hoặc đối tác của Viettel)
- ③ **Chiết khấu:**
  - Giảm thêm **10%** khi đăng ký theo **nhóm tính năng** (VD: Generator + Transformer)

