

Chương V. YÊU CẦU VỀ KỸ THUẬT

1. Giới thiệu chung về dự toán mua sắm, gói thầu:

- Tên gói thầu: Thuê dịch vụ chuyên ngữ và lồng phụ đề cho kênh Vietnam Today giai đoạn 2025-2026.

- Tên dự toán mua sắm: Thuê dịch vụ chuyên ngữ và lồng phụ đề cho kênh Vietnam Today giai đoạn 2025-2026.

- Nguồn vốn: Nguồn NSNN bổ sung năm 2025 thực hiện theo Quyết định số 1115/QĐ-THVN ngày 28/8/2025 của Tổng Giám đốc Đài THVN và dự kiến nguồn NSNN năm 2026.

- Chủ đầu tư: Trung tâm Công nghệ và Truyền dẫn phát sóng

- Địa điểm thực hiện: Trung tâm Công nghệ và Truyền dẫn phát sóng - Đài THVN, Số 43 Nguyễn Chí Thanh, Phường Giảng Võ, TP. Hà Nội.

- Nội dung gói thầu (Phạm vi cung cấp):

* Sử dụng ứng dụng AI để thực hiện tạo phụ đề tự động cho 8 thứ tiếng, bao gồm: tiếng Pháp, Nga, Trung, Tây Ban Nha, Ả rập, Hàn Quốc, Nhật Bản, Việt Nam.

* Chèn phụ đề tự động 8 thứ tiếng nêu trên vào luồng tín hiệu kênh Vietnam Today để cung cấp cho các nền tảng số phát sóng ở nước ngoài.

* Cung cấp 01 luồng tín hiệu kênh Vietnam Today với phụ đề gồm 8 thứ tiếng nêu trên để sử dụng phát sóng tại Việt Nam.

* Hệ thống thiết bị cung cấp dịch vụ được đặt tại VTV chỉ phục vụ riêng cho VTV; theo mô hình On premise, hệ thống hoạt động 24/7; có dự phòng 1+1.

- Thời gian thực hiện gói thầu: 10,5 tháng.

* Thiết lập dịch vụ: 15 ngày kể từ ngày ký hợp đồng.

* Cung cấp dịch vụ: 10 tháng kể từ ngày hoàn thành thiết lập dịch vụ.

2. Mục tiêu công việc:

Nhằm đảm bảo hiệu quả trong công tác quản lý, lưu trữ dữ liệu, trao đổi thông tin nội bộ, phối hợp công việc.

3. Yêu cầu kỹ thuật của gói thầu:

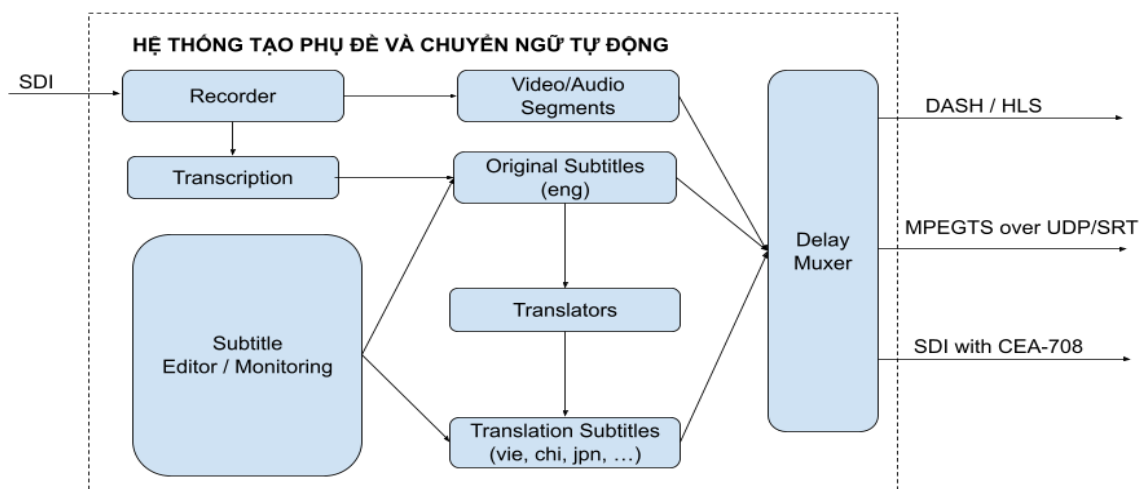
3.1. Yêu cầu chung:

Nhà thầu phải cung cấp đầy đủ các tài liệu kỹ thuật và các tài liệu khác minh họa để làm rõ phương án kỹ thuật, tính năng cũng như đáp ứng các chỉ tiêu kỹ thuật.

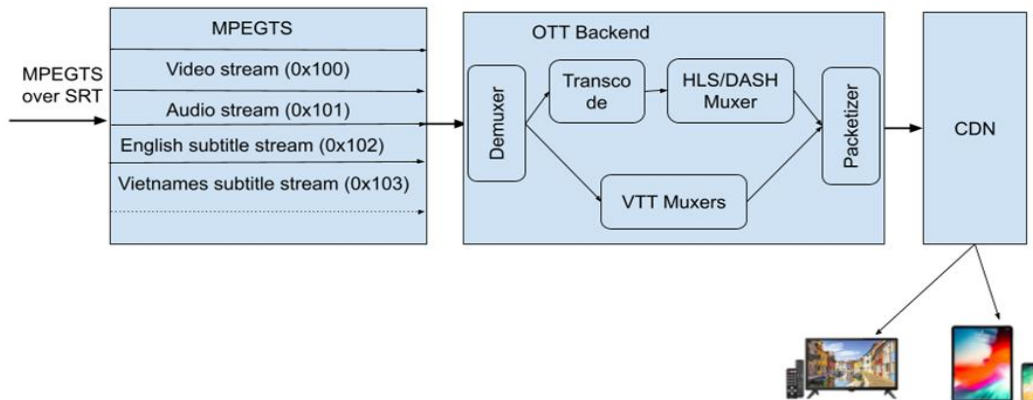
Trong phần tuyên bố đáp ứng của nhà thầu phải nêu rõ đã giải thích/dẫn chứng tại phần nào, mục nào, tài liệu nào của E-HSDT đáp ứng yêu cầu kỹ thuật gì trong E-HSMT để chủ đầu tư dễ dàng tham chiếu khi xem xét E-HSDT.

3.2. Yêu cầu kỹ thuật chi tiết:

a) Mô tả phương án xây dựng hệ thống phụ đề và chuyển ngữ:



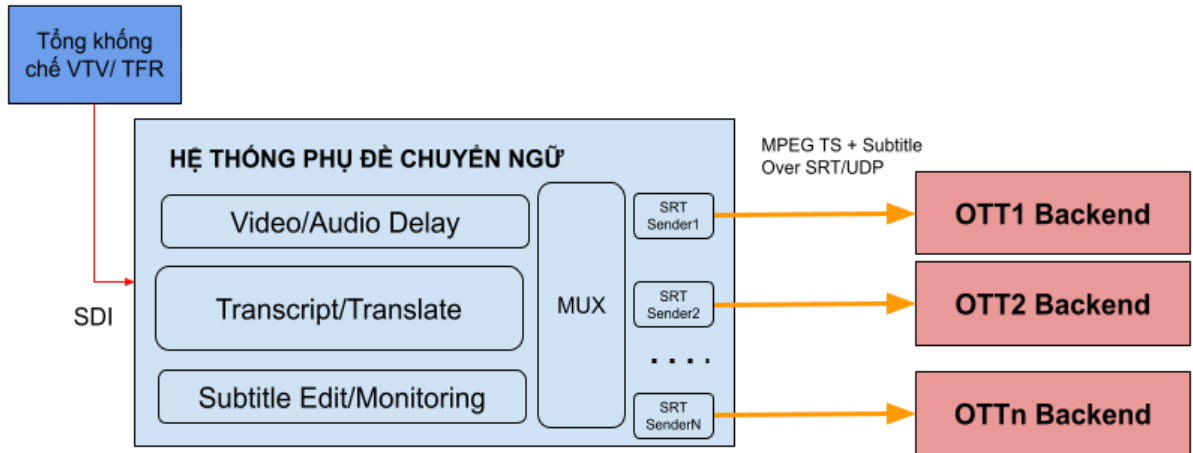
- Hệ thống nhận vào tín hiệu SDI từ Tổng không chế VTV/TFR.
- Tín hiệu SDI được đưa vào module Recorder để chuyển thành các file chứa các segment video/audio, đồng thời cung cấp audio cho module Transcription.
- Module Transcription sử dụng mô hình AI lắng nghe âm thanh được Recorder cung cấp để chuyển thành phụ đề gốc (ngôn ngữ được dùng trong tín hiệu đầu vào).
- Các module Translator (tương ứng với các ngôn ngữ cần dịch) nhận phụ đề gốc từ module Transcription, thực hiện chuyển ngữ thành các luồng phụ đề cho từng loại ngôn ngữ.
- Module Subtitle Editor/Monitoring cung cấp giao diện giám sát, chỉnh sửa phụ đề gốc và các phụ đề theo từng ngôn ngữ được dịch.
- Module Delay Muxer thực hiện ghép các luồng phụ đề (phụ đề gốc, phụ đề dịch ra các ngôn ngữ) vào luồng video/ audio để được 1 luồng hoàn chỉnh, luồng này có thể cung cấp dưới dạng luồng IP SRT hoặc MPEG DASH/ HLS hoặc chuyển lại thành tín hiệu SDI có chứa closed caption CEA-708 để phân phối.



Với luồng MPEG TS truyền tải qua SRT/UDP, luồng bao gồm các dòng video, audio và các dòng phụ đề, mỗi dòng có PID riêng (tự động hoặc đặt được trên Muxer) và có timestamp thống nhất để đồng bộ với nhau. Tại đầu nhận, tín hiệu này được đưa qua 1 bộ Demuxer để phân tách các dòng stream, các dòng video/audio được transcode theo format quy định của OTT sau đó đưa qua Muxer HLS (MPEG DASH) để đóng gói cho luồng media, đồng thời đưa qua các Muxer VTT để đóng gói cho phụ đề theo các ngôn ngữ, kết quả các files đầu ra cuối cùng được đẩy lên CDN để phân phối tới các thiết bị đầu cuối của người dùng, thông qua ứng dụng (app) hoặc trình duyệt.

- Đối với nhu cầu phân phối đa điểm qua SRT (video+audio+phụ đề chung trong 1 luồng SRT) sẽ cần bổ sung thêm các module truyền SRT cho từng điểm nhận (cấu hình phần cứng không thay đổi).

- Đối với nhu cầu phân phối đa điểm nhưng nhận với phương thức khác nhau (ví dụ: có điểm thì nhận SRT, có điểm thì nhận HLS, điểm thì nhận SDI...) sẽ cần bổ sung thêm các bộ phân phối tương ứng (tích hợp hệ thống), tùy thuộc vào số lượng điểm nhận và phương thức nhận mà có thể phải bổ sung thêm phần cứng. Nhà thầu có trách nhiệm cung cấp phần cứng đủ năng lực đáp ứng yêu cầu phân phối của Đài THVN.



*** Mô hình trí tuệ nhân tạo (AI):**

- Hệ thống sử dụng các mô hình AI được fine-tuning lại cho nhu cầu hệ thống.
- Mô hình có thể được tinh chỉnh theo bộ dữ liệu cho các lĩnh vực khác nhau theo yêu cầu để nâng cao độ chính xác. Bên cạnh đó, có thể sử dụng mô hình bổ sung được tinh chỉnh lại trên tập dữ liệu nhỏ để học được các từ ngữ Tiếng Việt quan trọng, được kết hợp với bộ từ điển và mô hình xác suất để nâng cao khả năng nhận diện được các từ mong muốn, kết quả từ mô hình này kết hợp với mô hình chính để nâng cao độ chính xác của toàn bộ hệ thống.
- Tất cả các mô hình AI đều được triển khai on-premise, không cần kết nối ra ngoài internet.
- Nhà thầu phải chịu trách nhiệm về bản quyền, đảm bảo an toàn thông tin theo quy định pháp luật đối với mô hình AI sử dụng.

b) Yêu cầu kỹ thuật chi tiết:

Nhà thầu phải cung cấp dịch vụ phải đáp ứng các yêu cầu kỹ thuật tối thiểu như sau:

TT	Yêu cầu kỹ thuật
1	Yêu cầu chung và phạm vi dịch vụ:
	<ul style="list-style-type: none"> - Cung cấp dịch vụ chuyển ngữ và lồng phụ đề tự động sử dụng nền tảng trí tuệ nhân tạo (AI) và học máy. - Thực hiện dịch vụ chuyển ngữ và lồng phụ đề sang 08 ngôn ngữ, bao gồm tiếng Pháp, tiếng Trung, tiếng Nga, tiếng Tây Ban Nha, tiếng Nhật,

TT	Yêu cầu kỹ thuật
	<p>tiếng Hàn Quốc, tiếng Việt, tiếng Ả Rập cho kênh truyền hình Vietnam Today.</p> <ul style="list-style-type: none"> - Tiêu chuẩn tín hiệu và phát sóng: <ul style="list-style-type: none"> + Luồng tín hiệu được đóng gói từ VTV Head End, phụ đề được đóng gói gửi đồng thời đến các địa điểm tiếp nhận OTT kênh Vietnam Toaday, VTVGo tại Đài THVN, (phụ đề dạng closed caption được truyền tách biệt khỏi tín hiệu video). + Phù hợp với chất lượng phát sóng HDTV (1080p 25/ hoặc 30; Video Codecs: H.264/AVC; Audio Codecs: AAC, MP3). + Hệ thống phải hỗ trợ các giao thức truyền video phổ biến như SRT, HLS (HTTP Live Streaming) và MPEG-DASH. + Tín hiệu Video/Audio và các phụ đề phải đóng trong cùng 1 luồng SRT truyền tới hệ thống nhận. + Hệ thống phải hỗ trợ đầu ra SDI có Closed Captions CEA-708.
2	Yêu cầu về hiệu suất và độ chính xác của AI:
	<p>Hệ thống phải có các chế độ hoạt động và độ chính xác của AI tự động tạo phụ đề từ âm thanh như sau:</p> <ul style="list-style-type: none"> + Chế độ độ trễ thấp với các chương trình trực tiếp: Đạt độ chính xác trên 90% với độ trễ tối đa là 5 giây. + Chế độ độ chính xác cao: Đạt độ chính xác trên 95% với độ trễ là 60 giây + Độ chính xác tối ưu: Có thể đạt tới 99% đối với các chương trình đã có sẵn phụ đề tiếng Anh/Việt hoặc với các ngôn ngữ đã có đủ dữ liệu huấn luyện. + Hệ thống phải có khả năng làm chậm hình ảnh (delay video) để đảm bảo phụ đề hoặc giọng nói được đồng bộ cùng với video. <p>Hệ thống AI phải có khả năng xử lý chính xác các ngữ cảnh và thuật ngữ nhạy cảm về chính trị hoặc địa lý (ví dụ: tên các đảo, địa danh chuẩn hóa, tên và chức danh chính trị gia,...)</p>
3	Yêu cầu về tùy chỉnh, giám sát và khả năng tự học:
	<ul style="list-style-type: none"> + Dịch vụ bao gồm bộ công cụ cho phép cấu hình độ trễ tín hiệu, cho phép các chuyên gia phụ đề (captioners) giám sát và chỉnh sửa trực tiếp kết quả do AI tạo ra theo thời gian thực.

TT	Yêu cầu kỹ thuật
	<ul style="list-style-type: none"> + Hệ thống cho phép đồng thời nhiều người cùng phối hợp biên tập phụ đề tin hiệu trực tiếp của một ngôn ngữ tại một thời điểm để rút ngắn thời gian biên tập. Bất kỳ thay đổi dữ liệu phụ đề bởi một chuyên gia phụ đề sẽ ngay lập tức được tự động đồng bộ tới giao diện chỉnh sửa phụ đề của các chuyên gia phụ đề khác. + Hệ thống phải có khả năng bật/ tắt phụ đề tại đầu ra bởi người giám sát. + Hệ thống phải cho phép trích xuất thành các file phụ đề trong 1 khoảng thời gian. + Hệ thống phải có khả năng "học" từ các kho dữ liệu chuyên biệt để cải thiện độ chính xác. + Hệ thống có thể được "huấn luyện" (train) riêng cho các lĩnh vực đặc thù như thể thao, tài chính, chính trị, vị trí địa lý, địa danh để nhận dạng thuật ngữ tốt hơn - Quản lý "từ điển tùy chỉnh" (Custom Dictionary): + Cung cấp bộ công cụ cho phép Biên tập viên VTV chủ động cập nhật, kiểm soát các từ ngữ nhạy cảm hoặc tên riêng theo yêu cầu. + Biên tập viên có thể bổ sung danh sách từ vựng, tên riêng của khách mời, địa danh, thuật ngữ... trước khi chương trình diễn ra để AI " nạp" thông tin và tăng độ chính xác. + Hệ thống có chức năng cho phép nạp phụ đề có sẵn của chương trình để sử dụng, cho phép người dùng đồng bộ thời gian lên phụ đề trên hệ thống giám sát cho cả file.
4	Yêu cầu về định dạng và hiển thị phụ đề:
	<ul style="list-style-type: none"> - Các định dạng file phụ đề cần hỗ trợ: .srt, .vtt, .ass, .cip. - Ngoài ra hệ thống phải hỗ trợ xuất phụ đề theo chuẩn EBU-TT-D (ETSI TS 103 285); đây là định dạng dựa trên TTML, tiêu chuẩn cho nhiều nền tảng OTT tại châu Âu. - Tùy chọn hiển thị: + Hỗ trợ đầy đủ các ký tự Unicode để hiển thị chính xác 08 ngôn ngữ mục tiêu và các ký hiệu đặc biệt. + Người xem tại đầu cuối có thể tùy chọn ngôn ngữ phụ đề muốn xem. Hệ thống cho phép hiển thị theo mức độ ưu tiên; nếu ngôn ngữ ưu tiên không có sẵn, hệ thống sẽ tự động hiển thị ngôn ngữ đầu tiên có trong danh sách.

TT	Yêu cầu kỹ thuật
5	Yêu cầu về hoạt động
	<ul style="list-style-type: none"> - Hệ thống thiết bị cung cấp dịch vụ được đặt tại VTV; theo mô hình On premise. - Hệ thống hoạt động 24/7; có dự phòng 1+1.
6	Thời gian thực hiện thiết lập dịch vụ: trong vòng 15 ngày kể từ ngày kí hợp đồng.
7	Thời gian cung cấp dịch vụ: 10 tháng kể từ ngày hoàn thành thiết lập dịch vụ.

*** Cách thức đánh giá độ chính xác tạo phụ đề:**

Nhà thầu phải cung cấp hồ sơ, công cụ, ... chứng minh mức độ đáp ứng độ chính xác tạo phụ đề theo như tuyên bố trong Hồ sơ dự thầu.

Chủ đầu tư có thể sử dụng một hoặc đồng thời nhiều phương pháp đánh giá phổ biến như: đánh giá bằng phần mềm và do con người đánh giá. Cụ thể như sau:

- Đánh giá tự động bằng phần mềm (định lượng): phổ biến thường dùng là đánh giá theo WER - Word Error Rate - số lượng từ sai. khi đó mô hình có WER càng thấp thì chất lượng càng cao, các mô hình có WER < 10% được coi là có chất lượng rất tốt.

Theo đó, để chấm điểm WER, cần cung cấp đầu vào văn bản chuẩn tương ứng với nguồn audio dùng để đánh giá, sau đó cho audio chạy qua hệ thống, kết quả thu được đầu ra văn bản do AI nghe được, đem so sánh với văn bản đầu vào để tính điểm WER:

$$WER = (S + D + I) / N$$

Với:

- + S: từ bị thay
- + D: từ bị mất
- + I: từ bị thêm sai
- + N: tổng số từ trong văn bản gốc.

Theo phương pháp đánh giá này, VTV sẽ cung cấp các văn bản đúng với nguồn media dùng để đánh giá, đưa tín hiệu này vào hệ thống, sau khi hết tín hiệu, hệ thống xuất lại nội dung nghe được và chấm điểm kết quả này.

Ngoài ra, có thể sử dụng các bộ dữ liệu chuẩn được cộng đồng đóng góp cho việc chấm điểm (benchmark datasets) như:

+ LibriSpeech

+ Tedlium

Để chấm điểm WER theo phương pháp này, có thể sử dụng công cụ nguồn mở jiwer chạy trên python hoặc tương đương, cung cấp văn bản nguồn và văn bản từ AI, công cụ sẽ tính toán và đưa ra điểm WER.

- Phương pháp Con người đánh giá (dựa trên cảm nhận): Phương pháp này sử dụng đánh giá chủ quan từ các chuyên gia, dựa trên việc so sánh các tiêu chí gồm:

- Độ chính xác ngữ nghĩa
- Sự hợp lý trong ngắt câu
- Sử dụng hợp lý dấu câu
- Xử lý tên riêng
- Nhận dạng chính xác ngữ chuyên ngành.

c) Phương án cài đặt, đấu nối, vận hành hệ thống

*** Phương án cài đặt, đấu nối:**

Nhà thầu có trách nhiệm xây dựng phương án, triển khai tích hợp hệ thống và vận hành, khai thác, bảo trì hệ thống thiết bị đặt tại Đài THVN trong thời gian cung cấp dịch vụ. Phương án triển khai phải được nêu rõ trong hồ sơ dự thầu. Một số yêu cầu chính như sau:

- Hệ thống tạo phụ đề và chuyên ngữ gồm 1 hoặc nhiều máy chủ (tùy theo cấu hình máy chủ, số lượng ngôn ngữ cần dịch) được kết nối mạng nội bộ với nhau.

- Hệ thống được cung cấp tín hiệu SDI đầu vào, từ Tổng khống chế/TFR.

- Hệ thống sẽ được đặt tại DC (tầng 7) kết nối vào mạng khai thác, cho phép ban kiểm duyệt truy cập (theo phân quyền) để kiểm soát nội dung phụ đề tự tạo và phụ đề chuyên ngữ.

- Tín hiệu đầu ra được đẩy qua giao thức SRT đến hệ thống OTT VTVgo, Vietnam Today hoặc đến các điểm Headend khác tại VTV (để phân phối ra các nước).

Bên cạnh đó, luồng tín hiệu ra cũng được đẩy lên hệ thống kiểm duyệt phát sóng của Ban Truyền hình đối ngoại phục vụ việc lưu trữ.

- Hệ thống được thiết kế có khả năng chạy dự phòng 1+1 (Cho phép chạy 2 server đồng thời xử lý cùng 1 công việc, đảm bảo hoạt động liên tục ngay cả khi 1 server bị lỗi).

*** Phương án vận hành:**

Nhà thầu có trách nhiệm xây dựng quy trình vận hành hệ thống, triển khai vận hành, sửa chữa, bảo trì hệ thống trong thời gian cung cấp dịch vụ. Một số điểm chính của phương án vận hành như sau:

- Tín hiệu SDI được cấp vào đầu vào của Hệ thống.
- Hệ thống xử lý và tạo phụ đề gốc (tiếng Anh), các phụ đề chuyển ngữ (tiếng Việt, tiếng Nhật, tiếng Trung, ...).
- Người kiểm duyệt truy cập trang quản lý, giám sát nội dung phụ đề được tạo ra và chỉnh sửa nếu nội dung chưa chính xác. Các nội dung chỉnh sửa có thể được tùy chọn cập nhật vào “từ điển tùy chỉnh” để huấn luyện AI.

3.3. Các yêu cầu khác:

+ Nhà thầu có cam kết dịch vụ cung cấp tuân thủ các quy định liên quan về bản quyền, được phép sử dụng hợp pháp của Việt Nam. Đồng thời tuân thủ các quy định về an toàn thông tin, an ninh mạng theo quy định của pháp luật, của Đài THVN.

+ Nhà thầu phải cập nhật phần mềm, đào tạo AI để cải thiện độ chính xác, độ trễ của dịch vụ cung cấp. Cam kết nâng cấp phần mềm, phần cứng để đáp ứng dịch vụ chuyển ngữ và lồng phụ đề cho kênh Vietnam Today đáp ứng việc phân phối, phát sóng trên các hạ tầng kỹ thuật trong nước và ngoài nước theo yêu cầu của Đài THVN.

+ Nhà thầu cam kết bảo mật thông tin do chủ đầu tư cung cấp. Trường hợp phát hiện lộ, lọt thông tin mà nguyên nhân từ phía nhà thầu, nhà thầu chịu mọi trách nhiệm trước pháp luật; cam kết tuân thủ các quy định về an toàn thông tin của Đài THVN trong quá trình triển khai, thực hiện gói thầu này. Nhà thầu phải đảm bảo rằng (có cam kết) thiết bị, phần mềm cung cấp trong dịch vụ chào thầu không bị nhà sản xuất/nhà thầu hoặc bên thứ ba cố ý cài các phần mềm gián điệp hoặc phần mềm có mục đích xấu như: lấy cắp thông tin khách hàng, truy cập trái phép,.. Trường hợp bên mời thầu phát hiện được những vấn đề nêu trên, bên mời thầu có quyền tạm dừng, hủy dịch vụ và nhà thầu phải chịu mọi rủi ro, chi phí liên quan, đồng thời chịu trách nhiệm trước pháp luật.

+ Nhà thầu có cam kết đảm bảo không gây ảnh hưởng đến hệ thống hiện có trong thời gian triển khai, cung cấp dịch vụ. Có giải pháp dự phòng, bypass hệ thống cung cấp dịch vụ khi có sự cố, đảm bảo không gián đoạn việc phát sóng kênh chương trình Vietnam Today.

Bàn giao các tài liệu vận hành cho Đài THVN gồm:

- Tài liệu khảo sát.
- Tài liệu thiết kế, tích hợp hệ thống.

- Tài liệu cấu hình.
- Tài liệu kiểm thử, từ điển, dữ liệu đào tạo AI.
- Tài liệu hướng dẫn quản trị, vận hành hệ thống.
- Tài liệu hướng dẫn xử lý sự cố.

4. Giải pháp và phương pháp luận:

Nhà thầu chuẩn bị đề xuất giải pháp, phương pháp luận tổng quát thực hiện dịch vụ theo các nội dung quy định tại Chương này, gồm các phần như sau:

1. Giải pháp và phương pháp luận;
2. Kế hoạch công tác.

5. Quy định về kiểm tra, nghiệm thu sản phẩm:

- Công tác kiểm tra, nghiệm thu được thực hiện theo quy định pháp luật hiện hành, hồ sơ mời thầu, hồ sơ dự thầu và yêu cầu của Chủ đầu tư.

- Nhà thầu có trách nhiệm cung cấp đầy đủ giải pháp, sản phẩm, hồ sơ, tài liệu, dữ liệu để chứng minh tính đáp ứng các yêu cầu kỹ thuật, phục vụ nghiệm thu và phối hợp thực hiện kiểm thử theo các tiêu chí chất lượng đã được phê duyệt.

- Sản phẩm chỉ được nghiệm thu khi đáp ứng đầy đủ yêu cầu kỹ thuật, chất lượng, khối lượng và kết quả kiểm tra, kiểm thử. Trường hợp không đạt, Nhà thầu phải khắc phục và thực hiện nghiệm thu lại mà không phát sinh chi phí.