

# Ngôn ngữ dân tộc thiểu số trong kỉ nguyên trí tuệ nhân tạo (AI): Bảo tồn và phát triển

Trần Thị Ngọc Anh

Email: tranngocanh01292@gmail.com  
Trưởng Đại học Sư phạm Quảng Tây  
Số 15 đường Dục Tài, Quận Thất Tinh,  
thành phố Quế Lâm, Quảng Tây, Trung Quốc

**TÓM TẮT:** Trong bối cảnh nền kinh tế thị trường và hội nhập quốc tế đang diễn ra mạnh mẽ, việc kế thừa, bảo tồn và phát triển các ngôn ngữ dân tộc thiểu số đang là một vấn đề thách thức. Điều đó đảm bảo hiệu quả của việc thực hiện phát triển kinh tế - xã hội bền vững của các dân tộc thiểu số. Hiện nay, với sự tiến bộ vượt bậc của khoa học công nghệ, sự phát triển nhanh chóng của internet tốc độ cao, chúng ta có thể tận dụng lợi thế của cuộc Cách mạng công nghiệp này để ngôn ngữ dân tộc thiểu số không bị mai một. Bài viết phân tích nguyên nhân ảnh hưởng đến sự sinh tồn các ngôn ngữ dân tộc thiểu số ở Việt Nam, các cơ hội từ kỉ nguyên trí tuệ nhân tạo (AI) mang lại cho các ngôn ngữ thiểu số có nguy cơ tuyệt chủng; việc thiết lập thư viện số về các ngôn ngữ dân tộc thiểu số đang có nguy cơ tuyệt chủng, các dịch vụ tài liệu kĩ thuật số hoá trên nền tảng tích hợp và chia sẻ dữ liệu để thúc đẩy sự phát triển của các ngôn ngữ dân tộc thiểu số đang có nguy cơ tuyệt chủng và thúc đẩy sự thịnh vượng và ổn định của các vùng dân tộc.

**TỪ KHÓA:** Trí tuệ nhân tạo, ngôn ngữ dân tộc thiểu số, bảo tồn, phát triển.

→ Nhận bài 27/7/2021 → Nhận bài đã chỉnh sửa 26/9/2021 → Duyệt đăng 15/01/2022.

**DOI:** <https://doi.org/10.15625/2615-8957/12220118>

## 1. Đặt vấn đề

Với công nghệ ngày càng tiên tiến, trí tuệ nhân tạo (AI) đã dần đi vào công việc và cuộc sống của chúng ta, đồng thời nó cũng ảnh hưởng đến phương thức giao tiếp của chúng ta, đặc biệt là trong việc sử dụng ngôn ngữ dân tộc. Trong thời đại trí tuệ nhân tạo, làm thế nào để bảo vệ ngôn ngữ dân tộc? Đây là một câu hỏi đáng để chúng ta suy nghĩ. Sự phát triển và thay đổi của ngôn ngữ luôn gắn liền chặt chẽ đến sự phát triển và thay đổi của xã hội loài người. Sự phát triển xã hội là điều kiện cơ bản và nguyên nhân bên ngoài của sự phát triển ngôn ngữ, và ảnh hưởng lẫn nhau của các yếu tố khác nhau trong hệ thống ngôn ngữ là nguyên nhân bên trong của sự phát triển ngôn ngữ; con người sử dụng ngôn ngữ để tổ chức suy nghĩ, trao đổi tư tưởng và tình cảm, tổ chức sản xuất xã hội và thúc đẩy lịch sử tiến bộ. Một khi ngôn ngữ mất đi, nền văn hóa độc đáo mà nó mang theo sẽ biến mất trong lịch sử. Để bảo tồn những ngôn ngữ đang dần bị mai một, các quốc gia đang chạy đua với thời gian và sử dụng công nghệ hiện đại, đặc biệt là công nghệ trí tuệ nhân tạo, để lưu giữ và bảo tồn ngôn ngữ độc đáo của họ. Sử dụng công nghệ trí tuệ nhân tạo mang đến những khả năng mới để bảo vệ ngôn ngữ đang gặp nguy cơ tiêu vong. So với các phương pháp truyền thống, công nghệ trí tuệ nhân tạo không chỉ cải thiện tốc độ và khả năng lưu trữ ngôn ngữ, mà còn góp phần cải tiến cơ sở dữ liệu ngôn ngữ qua học tập tương tác.

## 2. Nội dung nghiên cứu

### 2.1. Tổng quan về trí tuệ nhân tạo

Trí tuệ nhân tạo hay trí thông minh nhân tạo (Artificial intelligence - viết tắt là AI) là một ngành thuộc lĩnh vực khoa học máy tính (Computer science), là trí tuệ do con người lập trình tạo nên với mục tiêu giúp máy tính có thể tự động hóa các hành vi thông minh như con người. Các quá trình này bao gồm: học tập (thu thập thông tin và các quy tắc sử dụng thông tin), lập luận (sử dụng các quy tắc để đạt kết luận gần đúng hoặc xác định) và tự sửa lỗi. Các ứng dụng của AI thường được nhắc đến là hệ thống chuyên gia, nhận dạng giọng nói và thị giác máy tính (nhận diện khuôn mặt, vật thể hoặc chữ viết).

Hiện nay, có rất nhiều các loại công nghệ AI phổ biến trong lĩnh vực khoa học máy tính và nổi bật nhất là 4 loại sau đây:

- Loại 1: Máy phản ứng (Reactive machines). Các hệ thống AI này không có bộ nhớ và làm một nhiệm vụ cụ thể. Một ví dụ là Deep Blue, chương trình cờ vua của IBM đã đánh bại Garry Kasparov vào những năm 1990. Deep Blue có thể xác định các quân cờ trên bàn cờ và đưa ra dự đoán nhưng vì nó không có bộ nhớ nên nó không thể sử dụng các kinh nghiệm trong quá khứ để thông báo cho những người chơi mới trong tương lai.

- Loại 2: Bộ nhớ hạn chế (Limited memory). Các hệ thống AI này có bộ nhớ, vì vậy chúng có thể sử dụng các kinh nghiệm trong quá khứ để thông báo các quyết định trong tương lai. Một số chức năng ra quyết định trong xe tự lái được thiết kế theo cách này.

- Loại 3: Theory of mind. Lí thuyết của tâm trí là một thuật ngữ tâm lí học. Khi áp dụng vào AI, điều đó có nghĩa là hệ thống sẽ hiểu cảm xúc. Loại AI này sẽ có thể suy ra ý định và dự đoán hành vi khi nó được hoàn thiện.

- Loại 4: Tự nhận thức (Self-awareness). Trong thể loại này, các hệ thống AI có ý thức về bản thân, giúp chúng có ý thức. Máy móc tự nhận thức hiểu tình trạng hiện tại của chính họ. Loại AI này chưa tồn tại.

## 2.2. Các ứng dụng trí tuệ nhân tạo AI trong việc bảo vệ và phát triển các ngôn ngữ thiểu số

Một số ngôn ngữ đã tuyệt chủng trong một thời gian dài và vẫn còn lưu giữ, ẩn trong các văn bản vẫn chưa được giải mã, những thông tin vô giá về xã hội mà những người nói thời đó sinh sống. Các chuyên gia đã sử dụng các kĩ thuật hiện đại để hỗ trợ việc dịch thuật của họ, chẳng hạn như kĩ thuật hình ảnh trong trường hợp các văn bản khắc, nơi độ sâu của bản khắc ảnh hưởng đến ý nghĩa của văn bản. Nhưng một công nghệ dựa trên AI mới hiện giúp các chuyên gia giải mã các ngôn ngữ đã tuyệt chủng, giúp khôi phục một số nền văn hóa đã mất từ lâu.

Phương pháp chính xác và hiệu quả này đã cho phép dịch 2/3 ngôn ngữ đã tuyệt chủng, Linear B. Gần đây, trí tuệ nhân tạo DeepMind của Google cũng đã chứng minh được hiệu quả hơn con người trong việc dịch tiếng Hi Lạp cổ đại. Các phương pháp dựa trên AI khác, yêu cầu ít điểm dữ liệu hơn, có thể giúp người dịch giải mã các ngôn ngữ đã tuyệt chủng, ngay cả khi họ không có nhiều văn bản theo ý mình. Mỗi quan tâm thậm chí còn mạnh mẽ hơn đối với các ngôn ngữ hầu như vẫn chưa được biết đến, chẳng hạn như chữ hình nêm, trong đó chỉ có khoảng mười phần trăm văn bản được dịch cho đến nay. Các chương trình nghiên cứu đa quốc gia nhằm mục đích dịch hàng nghìn văn bản sang ngôn ngữ này với sự trợ giúp của AI.

Sự tham gia của các nhà khoa học trong lĩnh vực AI được áp dụng cho các ngôn ngữ cho thấy sự sẵn sàng sử dụng các kĩ thuật tiên tiến nhất hiện có của nhân loại để giải mã một quá khứ chưa được biết đến và bảo tồn sự phong phú và đa dạng mà các ngôn ngữ mang lại. Công nghệ này có thể mang lại những lợi ích xã hội và văn hóa mạnh mẽ và cung cấp cho chúng ta những chìa khóa để xây dựng tương lai của chúng ta từ những bài học của quá khứ.

## 2.3. Những lí do cho sự nguy cấp của các ngôn ngữ thiểu số

Theo thống kê chính thức năm 2019, “Có 32/53 dân tộc thiểu số (DTTS) có chữ viết riêng của dân tộc mình. Tỷ lệ người DTTS từ 15 tuổi trở lên biết đọc biết viết chữ của dân tộc mình còn thấp (15,9%). Tỷ lệ này cao nhất là ở dân tộc Ê Đê (38,8%), tiếp đến là các dân tộc

Ba Na (31,7%), Hoa (31,4%), thấp nhất là ở dân tộc Co, Lự (0,8%). Tỷ lệ nam giới biết đọc, biết viết chữ của dân tộc mình cao hơn nữ giới (17,5% so với 14,2%); ở khu vực thành thị cao hơn khu vực nông thôn (26,2% so với 14,4%)” [1].

### 2.3.1. Dân số sử dụng ngôn ngữ mẹ đẻ ngày càng giảm, thậm chí có nguy cơ tuyệt chủng

Theo kết quả điều tra dân số năm 2019, Việt Nam có 54 dân tộc, tổng số dân là 96,208,984 người. Trong số 54 dân tộc, dân tộc Kinh có dân số 8,085,826 người, chiếm 85,3% tổng dân số cả nước, là dân tộc đa số. Với dân số 14,123,158 người, chiếm 14,7% tổng dân số, 53 dân tộc còn lại là các DTTS [1]. Một số ngôn ngữ của đồng bào DTTS đứng trước nguy cơ mai một, trong đó có những ngôn ngữ gần như mất hoàn toàn như: tiếng Cơ Lao Đỏ ở Trùng Sán, Hà Giang, tiếng Ó Đu ở Con Công, Nghệ An, tiếng Tu Dí (Bố Y) ở Mường Khương, Lào Cai... Một số ngôn ngữ hiện chỉ còn rất ít người sử dụng, đó là tiếng Pu Páo, Cơ Lao Trắng, La Chí ở Hà Giang, tiếng Rục, Mày, Sách, Arem ở Quảng Bình [2].

Hiện nay, hầu hết các ngôn ngữ của các DTTS có nguy cơ tuyệt chủng ở nước ta không có chữ viết, một số đã mất chức năng giao tiếp hàng ngày và chỉ có một số ít người cao tuổi biết đến. Chúng chủ yếu được thể hiện bằng ngôn ngữ truyền miệng như những câu ca dao, truyền thuyết được kể lại. Nếu một ngôn ngữ biến mất, văn hóa chứa đựng ngôn ngữ ấy sẽ không còn. Nếu chúng ta không có biện pháp bảo vệ những ngôn ngữ đang suy yếu kịp thời, chúng ta sẽ mất đi một phần di sản văn hóa của nhân loại. Ngôn ngữ là phương thức biểu đạt chủ yếu trong giao tiếp của con người, nó cũng là biểu tượng bản sắc văn hóa dân tộc, đồng thời là ngôi nhà máu thịt và tinh thần của một dân tộc. Vì vậy, việc bảo vệ ngôn ngữ dân tộc (tiếng mẹ đẻ), đặc biệt là tiếng DTTS trở thành một nhiệm vụ quan trọng và cấp bách.

Các bản ghi của ngôn ngữ có nguy cơ tuyệt chủng chủ yếu dựa vào con người nhưng cấu trúc của chúng rất phức tạp và có thể thay đổi được. Việc chỉ dựa vào nguồn nhân lực thủ công để tóm tắt những thay đổi hoàn chỉnh về ngữ âm, phân tích ngữ nghĩa và phân tích ngữ pháp là không thực tế. Ngoài ra, cũng có nhiều khó khăn cụ thể trong việc thiếu chuẩn hóa việc thu thập và phiên âm văn bản của ngôn ngữ đang nghiên cứu.

### 2.3.2. Tiếng Anh phát triển và các ngôn ngữ dân tộc thiểu số bị ảnh hưởng ở một mức độ nhất định

Việt Nam là một đất nước có 54 dân tộc anh em, mỗi dân tộc đều có ngôn ngữ, bản sắc văn hóa riêng, góp phần tạo nên sự phong phú, đa dạng của nền văn hóa

Việt Nam. Trong bối cảnh nền kinh tế thị trường và hội nhập quốc tế đang diễn ra mạnh mẽ, việc Việt Nam gia nhập WTO, đối mặt với toàn cầu hóa đã kết nối chặt chẽ hơn với thế giới đặc biệt là xu hướng ra nước ngoài du lịch, du học và làm việc. Ví dụ: Nhiều trung tâm ngoại ngữ xuất hiện, một số trường tiểu học dạy tiếng Anh là ngôn ngữ chủ đạo và sau đó là tiếng Việt, tiếng mẹ đẻ nhường chỗ cho tiếng nước ngoài. Nhiều ngôn ngữ DTTS ở nước ta cũng đang đứng trước bờ vực tuyệt chủng, hiện tượng khử phương ngữ của các phương tiện truyền thông chính thống đang diễn ra nghiêm trọng, trẻ em có thể giao tiếp bằng tiếng Anh, tiếng phổ thông nhưng không thể giao tiếp được bằng ngôn ngữ của dân tộc mình. Tất cả mọi quốc gia và vùng lãnh thổ trên thế giới đều đang ra sức bảo vệ ngôn ngữ và văn hóa dân tộc nhưng chúng ta vẫn đang chạy theo sự cuồng nhiệt của ngoại ngữ, xem nhẹ quốc ngữ và phương ngữ.

### 2.3.3. Tác động của sự dịch chuyển dân số đối với ngôn ngữ dân tộc thiểu số

Với sự phát triển của quá trình đô thị hóa ở nước ta và tốc độ di dân tăng nhanh, cộng với sự dịch chuyển nhân khẩu tại các vùng DTTS, từ đó tạo nên một xã hội đa ngữ, phá vỡ dần tính bền chặt của các tiểu cộng đồng, nhất là các tiểu cộng đồng đơn dân tộc ngay tại các xóm bản. Hệ quả dẫn đến đối với ngôn ngữ là người ta phải tìm đến một ngôn ngữ chung tiếng Việt để giao tiếp. Nhờ đó, tiếng Việt trở thành ngôn ngữ chính, việc sử dụng tiếng DTTS có xu hướng giảm đi. Ngay cả ở những khu vực mật độ dân số đông, vì lí do kinh tế, nhiều bậc cha mẹ hướng con cái họ đến tiếng Việt và các ngoại ngữ để tìm kiếm việc làm. Việc ít tiếp xúc, sử dụng tiếng mẹ đẻ của đồng bào dân tộc và việc sử dụng tiếng Việt làm ngôn ngữ giao tiếp chính trong gia đình đã dẫn đến suy giảm khả năng sử dụng phương ngữ. Cơ cấu dân số thay đổi, dân số các vùng phương ngữ thu hẹp lại và dân số các vùng phương ngữ khác “đảo ngược”. Để thích nghi và hòa nhập với xã hội sở tại, đồng bào dân tộc tìm việc làm tại các thành phố lớn và vừa trên cả nước chủ động chọn tiếng Việt làm ngôn ngữ giao tiếp chính, địa bàn và phạm vi ngôn ngữ DTTS không ngừng bị thu hẹp.

Hơn thế nữa, xu hướng di cư của đồng bào DTTS tới những nơi mới ngày càng tăng lên khiến cho cường độ tiếp xúc của dân tộc này với dân tộc khác ngày càng cao, đặc biệt là sự tiếp xúc với người Việt (Kinh). Điều này cho thấy rằng, khuynh hướng người dân lựa chọn tiếng Việt ở vùng dân tộc sẽ ngày càng tăng.

### 2.4. Cơ hội từ kĩ thuật trí tuệ nhân tạo (AI) mang lại cho các ngôn ngữ thiểu số đang đứng trước nguy cơ tuyệt chủng

Sự phát triển của công nghệ thông tin và việc thiết lập

thế giới ảo đã phá vỡ giới hạn không gian và thời gian truyền thống, đồng thời làm cho việc giao lưu và trao đổi giữa các dân tộc trở nên thuận tiện hơn. Hơn nữa, do phát thanh, truyền hình, mạng internet và thông tin di động tiếp tục đi sâu vào đời sống hàng ngày của con người, mang lại nhiều tiện ích cho cuộc sống. Hiện tại, con người đang sử dụng công nghệ trí tuệ nhân tạo để nghiên cứu một cách có hệ thống các ngôn ngữ có nguy cơ tuyệt chủng và đã thu lại một số thành quả nhất định. Công nghệ trí tuệ nhân tạo có thể xây dựng một kho ngữ liệu hoàn chỉnh, và nội dung ghi âm bao gồm: xác định cấu trúc ngữ âm của ngôn ngữ, phân tích cấu trúc cú pháp của ngôn ngữ và sự tương ứng giữa các đơn vị nghĩa cơ bản. Có thể nói, trí tuệ nhân tạo AI đã mở ra “mùa xuân” cho sự bảo tồn và phát triển ngôn ngữ của các DTTS đang có nguy cơ tuyệt chủng.

*Thứ nhất*, trí tuệ nhân tạo (AI) cung cấp tính khả thi về mặt kĩ thuật cho việc phổ biến các ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng nhờ các đặc tính vô hình, tương tác và đa dạng của nó. Với sự hỗ trợ của Internet, các ngôn ngữ thiểu số có nguy cơ tuyệt chủng sẽ được chuyển đổi thành các phiên bản điện tử, sử dụng mạng máy tính và đài phát thanh, truyền hình, điện thoại di động và các phương tiện truyền thông khác để thúc đẩy việc sử dụng phương ngữ và thổ ngữ của riêng các dân tộc, địa phương và các vùng khác. Chính quyền các cấp cần đẩy mạnh việc xây dựng cơ sở hạ tầng Internet, các phương tiện truyền thông ở các vùng nông thôn, đặc biệt là các vùng sâu, vùng xa để thúc đẩy người dân trong việc trao đổi và sử dụng ngôn ngữ dân tộc mình. Việc này sẽ góp phần bảo vệ và kế thừa các ngôn ngữ dân tộc có nguy cơ tuyệt chủng.

*Thứ hai*, trí tuệ nhân tạo AI cung cấp hỗ trợ kĩ thuật cho việc bảo tồn và bảo vệ các ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng. Việc chỉ dựa vào nguồn nhân lực để tóm tắt những thay đổi hoàn chỉnh về ngữ âm, phân tích ngữ nghĩa và phân tích ngữ pháp là không thực tế. Ngoài ra, cũng có nhiều khó khăn cụ thể trong việc thiếu chuẩn hóa việc thu thập và phiên âm văn bản của ngôn ngữ đang nghiên cứu.

Sử dụng các công nghệ trí tuệ nhân tạo như nhận dạng giọng nói, phiên âm giọng nói và tổng hợp giọng nói, trọng âm và văn bản có thể được kết hợp tốt hơn để đạt được đầu vào bằng giọng nói bằng các ngôn ngữ có nguy cơ tuyệt chủng. Bằng cách này, ngay cả khi không có ngôn ngữ viết nó được chuyển đổi có điều kiện thành văn bản ngôn ngữ chính thống, có thể được ghi lại và truyền đi.

Có thể thực hiện nhân bản ngôn ngữ cho các ngôn ngữ mà cả ngôn ngữ viết và ngôn ngữ nói đều được sử dụng rộng rãi; cũng có thể thực hiện nhân bản ngôn ngữ cho các ngôn ngữ đã viết nhưng hiếm khi được sử dụng

và ngôn ngữ nói vẫn được sử dụng bình thường. Đối với các ngôn ngữ không có văn bản, lời nói của ngôn ngữ được nghiên cứu có thể được thu thập và phiên âm thành văn bản của ngôn ngữ chính trong các điều kiện. Sử dụng dữ liệu song song như vậy, công nghệ đầu cuối của trường trí tuệ nhân tạo có thể được sử dụng để thực hiện Việc chuyển đổi giữa các văn bản, tức là “chương trình Rosetta phiên âm” đang được các nhà khoa học Mỹ thực hiện, về mặt logic là khả thi.

*Thứ ba*, việc triển khai trí tuệ nhân tạo AI nâng cao ý thức bảo vệ ngôn ngữ của dân tộc. Việc thực hiện kế hoạch hành động trí tuệ nhân tạo AI đòi hỏi phải vượt qua các giới hạn về thời gian, không gian và dân tộc, thúc đẩy quyền tự do sử dụng ngôn ngữ mẹ đẻ của tất cả các nhóm dân tộc, đặc biệt là những dân tộc có dân số ít, đồng thời đây cũng là cơ hội tuyệt vời để quảng bá rộng rãi ngôn ngữ đang có nguy cơ mai một đến với nhiều người. Kết hợp trí tuệ nhân tạo AI với các ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng, thông qua các kênh truyền thông mới trên mạng Internet, có thể thúc đẩy đồng bào DTTS có ý thức hơn nữa trong việc bảo vệ và kế thừa văn hóa ngôn ngữ dân tộc của họ.

Hầu hết các ngôn ngữ thiểu số có nguy cơ tuyệt chủng đều được lưu giữ trong các chữ viết cổ, tất cả tinh hoa trí tuệ của các DTTS đều dựa vào đây để truyền lại cho con cháu, ngôn ngữ tuyệt chủng là một tổn thất nghiêm trọng. Các ngôn ngữ có nguy cơ tuyệt chủng dưới góc độ lịch sử xã hội phát triển ngôn ngữ không thể hồi sinh được, do đó, cần tận dụng lợi thế trí tuệ nhân tạo và Internet để tránh thất truyền các ngôn ngữ thiểu số.

### **2.5. Các biện pháp bảo vệ và kế thừa các ngôn ngữ thiểu số có nguy cơ tuyệt chủng trong kỉ nguyên trí tuệ nhân tạo**

Ngày nay, trí tuệ nhân tạo (AI) đã không còn xa lạ và bản thân ai cũng cảm nhận được tác động của nó đối với cuộc sống hàng ngày của chúng ta: nhận diện khuôn mặt, điều khiển bằng giọng nói, xe không người lái, nhà thông minh, ... Nhưng trên thực tế, sứ mệnh trí tuệ nhân tạo không chỉ đơn thuần là tạo điều kiện phát triển nhân loại mà còn bảo vệ nền văn minh của Trái đất.

#### **2.5.1. Thiết lập cơ sở dữ liệu kĩ thuật số về các ngôn ngữ dân tộc thiểu số nguy cấp**

Cuộc Cách mạng công nghiệp đã ảnh hưởng thay đổi cách sống của chúng ta và cung cấp hỗ trợ kĩ thuật mới cho việc thiết lập cơ sở dữ liệu kĩ thuật số về các ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng. Cơ sở dữ liệu kĩ thuật số chủ yếu bao gồm dữ liệu âm thanh và video bằng ngôn ngữ thiểu số. Sử dụng trí tuệ nhân tạo để thu thập cơ sở dữ liệu kĩ thuật số sinh thái ban đầu về các ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng

là một phương pháp khoa học và hiệu quả để bảo tồn các ngôn ngữ này. Ủy ban Ngôn ngữ quốc gia nên mở thêm các đề tài nghiên cứu ngôn ngữ thiểu số cho các trường đại học và viện nghiên cứu, tìm hiểu tình hình cụ thể của các ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng và yêu cầu chủ nhiệm dự án thu thập và sưu tầm các ngôn ngữ của một nhóm thiểu số có nguy cơ tuyệt chủng cao. Đồng thời, chính quyền các cấp cũng cần coi trọng việc kế thừa, bảo vệ tiếng dân tộc và xây dựng các biện pháp chiến lược tương ứng để tăng cường bảo vệ và lưu truyền.

Việc xây dựng cơ sở dữ liệu kĩ thuật số các ngôn ngữ thiểu số nguy cấp có thể sửa chữa các lỗi của hồ sơ giấy, có thể ghi lại tính toàn vẹn của thông tin ngôn ngữ thiểu số ở mức độ lớn nhất, ngăn ngừa mất thông tin dữ liệu cơ sở dữ liệu ngôn ngữ thiểu số đang có nguy cơ tuyệt chủng, đóng vai trò quan trọng trong việc bảo vệ cơ sở dữ liệu âm thanh ngôn ngữ thiểu số.

#### **2.5.2. Hiện thực công tác số hoá ngôn ngữ của các dân tộc thiểu số**

Để ngôn ngữ của các DTTS có thể tồn tại và bình đẳng với ngôn ngữ khác, tạo nên sự đa dạng văn hóa thì một trong các giải pháp không thể bỏ qua trong thời đại công nghệ 4.0 là công tác số hóa. Theo Tạ Văn Thông (2018) cùng với các biện pháp “nghiên cứu cơ bản về cấu trúc, tình hình xã hội ngôn ngữ học, cải tiến và xây dựng các hệ thống chữ viết; biên soạn các sách công cụ (sách giáo khoa, sách ngữ pháp, từ điển...); sưu tập các văn bản (vốn văn nghệ truyền thống; sáng tác mới...) và ghi bằng các ngôn ngữ có nguy cơ mai một; dạy và học các ngôn ngữ có nguy cơ tiêu vong và sử dụng chúng trên các phương tiện thông tin đại chúng” là việc “thu thập; lưu trữ để xây dựng ngân hàng dữ liệu” không chỉ “giúp cho người bản ngữ có ý thức bảo tồn, phát triển tiếng mẹ đẻ” mà còn hình thành một kho tài nguyên thông tin chân thực vô giá, giúp cho các thế hệ nghiên cứu hiện tại và sau này có tư liệu chính xác - trong mọi tình huống bất trắc có thể xảy ra. Đề cập đến “Các loại từ điển đối dịch cần có ở Việt Nam”, Tạ Văn Thông (2018) cho rằng: “Trong sự đa dạng này, không nên “bỏ quên” các ngôn ngữ DTTS đang đứng trước nguy cơ tiêu vong như Pà Thẻn, Pu Péo, La Ha, Bó Y, Cơ Lao, Co, Rơ Măm...” [3].

Để thực hiện số hoá ngôn ngữ của các DTTS đang có nguy cơ tuyệt chủng là số hóa nội dung của các ngôn ngữ thiểu số, số hóa các văn bản gốc bằng các ngôn ngữ thiểu số và thực hiện việc đọc kĩ thuật số trên nhiều phương tiện hơn. Sau khi tài liệu bằng tiếng DTTS được chuyển đổi thành ấn phẩm kĩ thuật số, chúng có thể được hiển thị trên điện thoại thông minh, máy tính bảng và các thiết bị đầu cuối khác. Ủy ban Ngôn ngữ Quốc gia và Ủy ban Dân tộc nên xây dựng các ứng

dụng (APP) bằng từng ngôn ngữ DTTS để cho đồng bào các DTTS tiện nắm bắt. Đây là một yêu cầu mới đối với việc xử lý thông tin kỹ thuật số bằng các ngôn ngữ thiểu số. Các ngôn ngữ thiểu số được đặt trên giao diện điện thoại thông minh, chữ viết tay và nhận dạng giọng nói được sử dụng để đóng vai trò của các ngôn ngữ thiểu số. Hiện tại, “công nghệ HTML5 và Android phải được sử dụng để xây dựng nền tảng kỹ thuật số cho các tài liệu tiếng DTTS và đây cũng là một chủ đề quan trọng cho việc thông tin hóa các tài liệu tiếng DTTS” [4]. Thông qua việc phát triển, truy cập và các ứng dụng (APP) khác nhau, việc tìm kiếm và truy vấn các ngôn ngữ thiểu số được thực hiện trên môi trường mạng để thúc đẩy ứng dụng trên phạm vi rộng và đa góc độ của các ngôn ngữ thiểu số.

Cách đây không lâu, nhà tương lai học người Mỹ, giám đốc điều hành Viện DaVinci, ông Thomas Frey đã đề xuất ý tưởng “kho lưu trữ ngôn ngữ toàn cầu”. Ông Thomas Frey đã sử dụng công nghệ AI để lưu trữ các ngôn ngữ dân tộc quốc gia khác nhau dưới dạng video, âm thanh và văn bản. Hiện tại, ông Frey kết hợp với các học giả khắp thế giới tham gia dự án ngôn ngữ có nguy cơ tuyệt chủng (Endangered Languages Project), được điều hành bởi Đại học Hawaii, đây là bước đầu tiên để tạo ra một kho lưu trữ ngôn ngữ toàn cầu. Nó thu thập tài nguyên và thông tin của các ngôn ngữ gặp nguy hiểm, đã thu thập dữ liệu từ gần 4000 ngôn ngữ, nhiều trong số đó được liệt kê trong danh sách các ngôn ngữ có nguy cơ tuyệt chủng [5].

Các nhà ngôn ngữ học Úc ở Trung tâm động thái ngôn ngữ ưu việt ARC (CoEDL) đã hợp tác với công ty Google của Mỹ, gần đây đã phát triển một nền tảng trí tuệ nhân tạo cho các ngôn ngữ có nguy cơ tuyệt chủng; nền tảng này sử dụng các mô hình trí tuệ nhân tạo để phiên âm các ngôn ngữ bản địa và đã lưu trữ hơn 40.000 giờ ghi âm. Điều này không chỉ tiết kiệm rất nhiều thời gian mà còn cải thiện độ chính xác của việc ghi âm. Hiện tại, họ đã thiết lập các mô hình trí tuệ nhân tạo cho hơn 20 ngôn ngữ bản địa của Úc [6].

Ở Trung Quốc, công ty ITFYTEK sử dụng công nghệ trí tuệ nhân tạo để tìm ra mối liên kết và cấu trúc giữa các ngữ nghĩa âm tiết ở các thang bậc khác nhau trong hệ thống ngôn ngữ đang có nguy cơ tuyệt chủng. Với công nghệ giọng nói thông minh tiên tiến, hệ thống giọng nói của công ty bao gồm nhiều ngôn ngữ và phương ngữ thiểu số đã được sử dụng thực tế trong các phương thức nhập liệu bằng giọng nói, phiên dịch, đánh giá và phát sóng tự động và đã đóng góp phần bảo trì sự sinh thái sinh thái của ngôn ngữ.

### 2.5.3. Thiết lập nền tảng trí tuệ nhân tạo chia sẻ ngôn ngữ thiểu số nguy cấp

Công nghệ trí tuệ nhân tạo đã tạo nên nhiều lợi ích

trong cuộc sống. Ngày càng có nhiều đồng bào DTTS sử dụng điện thoại di động giao diện ngôn ngữ mẹ đẻ để liên lạc và giao tiếp. Nếu công nghệ thông tin và công nghệ kỹ thuật số của nền tảng dữ liệu lớn AI được sử dụng đầy đủ trong việc bảo vệ và kế thừa các ngôn ngữ thiểu số có nguy cơ tuyệt chủng, một nền tảng dùng chung sẽ được thiết lập để phổ biến rộng rãi các ngôn ngữ thiểu số có nguy cơ tuyệt chủng, đồng thời làm sâu sắc thêm việc kế thừa và phát triển các ngôn ngữ thiểu số.

Trong kỉ nguyên trí tuệ nhân tạo, sự kết hợp hiệu quả giữa các cá nhân và phương tiện truyền thông cho phép mỗi cá nhân trở thành nhà xuất bản, người phát ngôn và người truyền đạt ngôn ngữ dân tộc. Điều này làm giảm thời gian và không gian giao lưu văn hóa và ngôn ngữ giữa các DTTS, đặc biệt đối với các vùng DTTS có vị trí xa xôi và thông tin kém phát triển. Ví dụ, sự ra đời của zalo, facebook đã phát huy hết tác dụng của chức năng giao tiếp này. Người gửi và người nhận thông tin có thể là mỗi người dùng và giao tiếp giữa hai bên được tăng cường thông qua phương tiện mới này. Việc xử lý thông tin của các ngôn ngữ thiểu số có thể được đảm bảo thông qua Internet, điện thoại thông minh và việc phổ biến văn hóa và ngôn ngữ dân tộc có thể được thực hiện. Việc thiết lập một nền tảng chia sẻ có thể sử dụng tốt hơn chức năng tìm kiếm các ngôn ngữ DTTS đang có nguy cơ tuyệt chủng. Thông tin về ngôn ngữ DTTS cần thiết cần được lựa chọn, tùy chỉnh, phổ biến và chấp nhận theo nhóm, đảm bảo hiệu quả việc trao đổi ngôn ngữ giữa các nhóm dân tộc và tối ưu hóa nguồn thông tin ngôn ngữ dân tộc, Mở rộng đối tượng cũng có lợi cho nghiên cứu liên ngành và giao tiếp trong thế giới học thuật.

### 3. Kết luận

Bảo vệ đa dạng ngôn ngữ là bảo vệ đa dạng văn hóa và bảo vệ đa dạng văn hóa là bảo vệ sự đa dạng của cộng đồng với một tương lai chung cho nhân loại. Có thể nói, ngôn ngữ không chỉ là một trong những đặc trưng quan trọng nhất của sự tồn tại của một dân tộc, mà còn là biểu tượng quan trọng của sự tiến bộ của con người, văn minh và xã hội hài hòa, ổn định. Nhờ vào lợi thế của công nghệ trí tuệ nhân tạo trong việc lưu trữ và học ngôn ngữ đã góp phần bảo tồn và lưu giữ ngôn ngữ có nguy cơ tuyệt chủng. Tuy nhiên, vẫn còn nhiều thử thách lớn để bảo vệ các ngôn ngữ qua công nghệ này. Ví dụ: Một số ngôn ngữ DTTS thường được truyền miệng không có kí tự, điều đó có thể thay đổi hoặc không thể diễn tả hoàn toàn ý nghĩa gốc của nó. Nhưng chúng ta có thể tin rằng, nhờ sự trợ giúp của công nghệ này, vấn đề bảo vệ ngôn ngữ có thể được giải quyết trong tương lai gần.

**Tài liệu tham khảo**

- [1] *Kết quả Điều tra thu thập thông tin về thực trạng kinh tế - xã hội của 53 dân tộc thiểu số năm 2019*, (2020), Tổng cục Thống kê, NXB Thống kê.
- [2] Tạ Văn Thông, (2018), *Biên soạn từ điển đối dịch trước nguy cơ mai một các ngôn ngữ ở Việt Nam*, Ki yếu Hội thảo quốc gia Nghiên cứu thực trạng và giải pháp bảo tồn ngôn ngữ dân tộc thiểu số có nguy cơ mai một ở Việt Nam hiện nay, Trường Đại học Sư phạm – Đại học Thái Nguyên, Thái Nguyên.
- [3] REN Ke, (2014), *Study on the HTML5 build Android minority literature digital platform doi:10.3969/j.issn.1003-4271.2014.01.26*.
- [4] F.Frey, (2012), *Creating a Global Language Archive* [https://www.academia.edu/2497650/Creating\\_a\\_Global\\_Language\\_Archive](https://www.academia.edu/2497650/Creating_a_Global_Language_Archive) FuturistSpeaker.com.
- [5] Stephen Ibaraki, (2018), *Turning To AI To Save Endangered Languages*. <https://www.forbes.com/sites/cognitiveworld/2018/11/23/turning-to-ai-to-save-endangered-languages/?sh=516b8f9f6f45>.
- [6] Timbigg, (2018), *The social robot that could help save indigenous languages* <https://www.smh.com.au/technology/the-social-robot-that-could-help-save-indigenous-languages-20180601-p4ziyj.html> The social robot that could help save indigenous languages.
- [7] Vương Toàn, (2018), *Khai thác tiện ích của công nghệ truyền thông nhằm bảo tồn ngôn ngữ dân tộc thiểu số có nguy cơ mai một ở nước ta*, Ki yếu Hội thảo khoa học quốc gia Nghiên cứu thực trạng và giải pháp bảo tồn những ngôn ngữ dân tộc thiểu số có nguy cơ mai một ở Việt Nam hiện nay, Trường Đại học Sư phạm Thái Nguyên, Trường Đại học Tây Bắc, Sơn La.
- [8] Vương Toàn, (2019), *Số hoá để bảo tồn bằng lưu trữ dữ liệu về các ngôn ngữ trước nguy cơ mai một: Từ trường hợp các ngôn ngữ Thái - Ka đai ở Việt Nam*, Ki yếu Hội thảo Ngôn ngữ dân tộc thiểu số có nguy cơ mai một ở Việt Nam hiện nay: Những vấn đề lí luận và thực tiễn, Trường Đại học Sư phạm - Đại học Thái Nguyên, Thái Nguyên.
- [9] Cai Zixing, (2016), *Artificial Intelligence: Principles & Applications*, NXB Đại học Thanh Hoa, Bắc Kinh.

**MINORITY LANGUAGES IN THE ERA OF ARTIFICIAL INTELLIGENCE (AI): CONSERVATION AND DEVELOPMENT**

**Tran Thi Ngoc Anh**

Email: tranngocanh01292@gmail.com  
Guangxi Normal University  
15 Yucai Road, Qixing District, Guilin, P.R.China

**ABSTRACT:** *In the context of the market economy and international integration, the inheritance, conservation and development of minority languages is a challenging issue. That ensures the effectiveness of the implementation of sustainable socio-economic development of ethnic minorities. With the progress of science and technology and the rapid development of high-speed internet, we can take advantages of this industrial revolution to protect the minority languages. The article analyzes the causes that affect the survival of minority languages in Vietnam, the opportunities in the era of artificial intelligence (AI) to endangered minority languages; the establishment of a digital library of endangered minority languages, integrated digital document services and data sharing to promote the development of minority languages which are threatened "as well as improve" the prosperity and stability of ethnic groups.*

**KEYWORDS:** Artificial intelligence, minority languages, conservation, development.