



# exploiter la puissance de l'IA en périphérie

Un entretien avec François de Rochebouët de STMicroelectronics

## Questions de l'équipe éditoriale d'Elektor

Découvrez comment les solutions innovantes d'IA pour les microcontrôleurs révolutionnent les systèmes embarqués et transforment les industries grâce à des prises de décision en temps réel et économies en énergie. François de Rochebouët (responsable marketing pour les solutions d'IA, chez STMicroelectronics) évoque le rôle essentiel de l'IA en périphérie (Edge AI) dans la construction d'un avenir durable.

**Elektor :** Merci François de nous avoir rejoints. Pourriez-vous commencer par vous présenter, et nous parler un peu de votre cheminement dans le domaine de l'IA et de l'innovation ?

**François de Rochebouët :** J'ai toujours été passionné par l'innovation et les limites du possible, en particulier dans le domaine des systèmes embarqués et de l'IA. En tant que cofondateur de Cartessim, nous avons été les premiers à proposer des solutions d'IA pour les microcontrôleurs. Aujourd'hui, chez STMicroelectronics, je dirige le marketing de l'IA pour les microcontrôleurs STM32. Ma vision est claire : l'IA en périphérie (Edge AI) est cruciale pour l'avenir. Nous avons besoin d'une IA à la périphérie, plus proche de l'endroit où les données sont générées, car les solutions basées sur l'informatique en cloud ne seront pas en mesure de gérer l'énorme quantité de données, générées par notre monde de plus en plus connecté. L'IA durable doit traiter les données localement, ce qui permet d'économiser de l'énergie et d'améliorer la réactivité.

The infographic is titled "Why edge AI matters" and features a yellow header bar with the text "Edge = where the SIGNAL becomes DATA". Below this, there is a large blue arrow pointing to the right. To the left of the arrow, there is a vertical list of five advantages of Edge AI, each with an icon and text:

- Ultra-low latency: Real-time applications
- Reduced data transmission: Generate meaningful information
- Enhanced privacy and security: No data sharing in the cloud
- Sustainable on energy: Low data, low power
- Lower cost of inference to enable a new class of operations

On the right side of the arrow, there is a close-up photograph of a person's fingers holding a small black STM32 microcontroller chip. The background of the slide is a blue grid with numbers and the ST logo at the bottom left.

Figure 1. Les avantages de l'IA en périphérie.

**Elektor :** Pourquoi l'**Edge AI** est-elle essentielle à la construction d'un avenir durable ?

**François :** La quantité de données que nous générions aujourd'hui est considérable, et ne va qu'augmenter de manière exponentielle. Le recours à l'IA en nuage (cloud) est synonyme de consommation d'énergie élevée, de problèmes de bande passante et de latence. L'**Edge AI** (**figure 1**) traite les données localement, à la source, réduisant ainsi la nécessité d'un transfert constant de données vers le cloud. Cela permet non seulement de réduire la consommation d'énergie, mais aussi de prendre des décisions plus rapides et plus efficaces. Pour assurer la durabilité, nous avons besoin de solutions d'intelligence artificielle capables de fonctionner dans des environnements à faible consommation d'énergie, et c'est là où les systèmes embarqués et les microcontrôleurs sont excellents.

**Elektor :** Quels sont les principaux défis auxquels les développeurs sont confrontés lorsqu'ils conçoivent des solutions d'IA en périphérie ?

**François :** L'un des principaux défis consiste à changer la façon dont les développeurs envisagent la résolution des problèmes. Dans les systèmes conventionnels, les développeurs s'appuient sur des capteurs pour obtenir des informations directes, mais ils changent d'approche avec l'IA. Elle nous permet de déduire des informations complexes à partir de données apparemment sans rapport. Cela nécessite un nouvel état d'esprit. Les développeurs sont également confrontés à des contraintes telles que la limitation de la puissance de traitement, de la mémoire et de l'efficacité énergétique. Ils doivent donc optimiser les modèles d'IA pour qu'ils s'adaptent à ces paramètres contraignants, tout en garantissant la précision et la fiabilité.

**Elektor :** Pouvez-vous donner des exemples de la façon dont l'IA embarquée a contribué à résoudre les problèmes des clients de manière créative ?

**François :** Les vélos à assistance électrique de Panasonic en sont un exemple. Au lieu d'utiliser des capteurs de pression traditionnels pour la surveillance des pneus, ils ont exploité les données de mesure de l'intensité du courant du moteur. En appliquant l'IA, ils ont créé un capteur de pression virtuel qui déduit la pression des pneus à partir des données existantes. Cela a permis de réduire la complexité et le coût du matériel tout en améliorant les performances. Un autre exemple est la détection d'anomalies dans les moteurs fonctionnant à des vitesses différentes. En utilisant l'IA, nous avons aidé nos clients à prévoir les défaillances à un stade avancé et à minimiser les temps d'arrêt, optimisant ainsi le calendrier de maintenances.

**Elektor :** Comment STMicroelectronics relève-t-elle le défi de concevoir des solutions d'IA en périphérie, avec des ressources limitées ?

**François :** Nous pensons que c'est la combinaison de ressources matérielles de haute performance et d'un écosystème logiciel adapté qui fait la différence. Sur le plan matériel, nous disposons des microcontrôleurs STM32 [1], leaders dans leur catégorie, qui offrent l'efficacité énergétique et les performances nécessaires à l'**Edge AI**. Côté logiciel, nous avons rassemblé nos outils et ressources dans la ST Edge AI Suite [2] (**figure 2**), qui rationalise le parcours du développeur, du concept au déploiement, et offre une large collec-

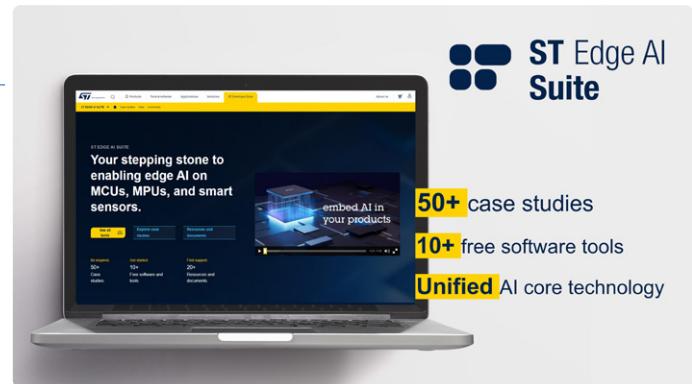


Figure 2. La ST Edge AI Suite, un ensemble d'outils permettant d'intégrer des fonctions d'intelligence artificielle dans les systèmes embarqués.

tion d'études de cas pour l'inspiration. De plus, la STM32 Developer Zone [3] offre un support, des tutoriels et des ressources complets, aidant les développeurs à créer des solutions innovantes avec les STM32. Cette combinaison de matériel et de logiciel permet aux développeurs de relever les défis et de créer de puissantes applications d'IA, quel que soit leur niveau de connaissances.

**Elektor :** Comment STMicroelectronics soutient-elle les développeurs ayant des niveaux de compétences différents dans le domaine de l'**Edge AI** ?

**François :** Chez ST, nous comprenons que les développeurs viennent d'horizons différents, c'est pourquoi nous avons adapté nos outils pour répondre à leurs besoins. Pour les nouveaux venus dans le domaine de l'IA, nous proposons NEAI Studio [4], une plateforme intuitive pour créer et déployer des modèles d'IA avec un minimum d'effort. NEAI Studio supporte une approche BYOD (Bring Your Own Device), avec une solution AutoML générant directement des bibliothèques, optimisées d'algorithmes d'apprentissage automatique les plus adaptés, en fonction des ensembles de données fournis par les utilisateurs.

Pour les utilisateurs avancés, STM32Cube.AI [5], ST Edge AI Core [6] et ST Edge AI Developer Cloud [7] fournissent des solutions d'optimisation et de benchmarking avancées. Les développeurs peuvent également bénéficier de scripts pour réentraîner des modèles à partir du « zoo de modèles » [8] ou de leurs propres modèles, en adoptant l'approche BYOM (Bring Your Own Model). Ce support complet permet aux développeurs d'intégrer leurs modèles pré-entraînés dans l'écosystème, en vue d'une optimisation et d'un déploiement plus poussés. De cette manière, que vous soyez un maker, un ingénieur chevronné dans les systèmes embarqués ou un data scientist, nous avons des outils qui correspondent à votre niveau d'expertise, et vous permettent de tirer le maximum de votre productivité et de votre innovation.

**Elektor :** Quel est le rôle du matériel dans les cas d'utilisation plus complexes de l'IA, comme la vision par ordinateur ?

**François :** Le matériel joue un rôle important, d'autant plus que les cas d'utilisation de l'IA deviennent de plus en plus complexes. Par exemple, la vision par ordinateur nécessite une puissance de calcul et une mémoire importantes. C'est là qu'intervient le matériel spécialisé, comme notre ST Neural-ART accelerator, une unité de traitement neuronal (NPU) conçue pour les microcontrôleurs STM32. Il est sans précédent en termes de puissance de calcul et d'efficacité énergétique, permettant aux développeurs d'implémenter des algorithmes d'IA plus exigeants, tels que la détection d'objets en temps réel, tout en conservant une faible consommation d'énergie.

# From DMIPS to TOPS, the paradigm shift Opening a new range of embedded AI applications

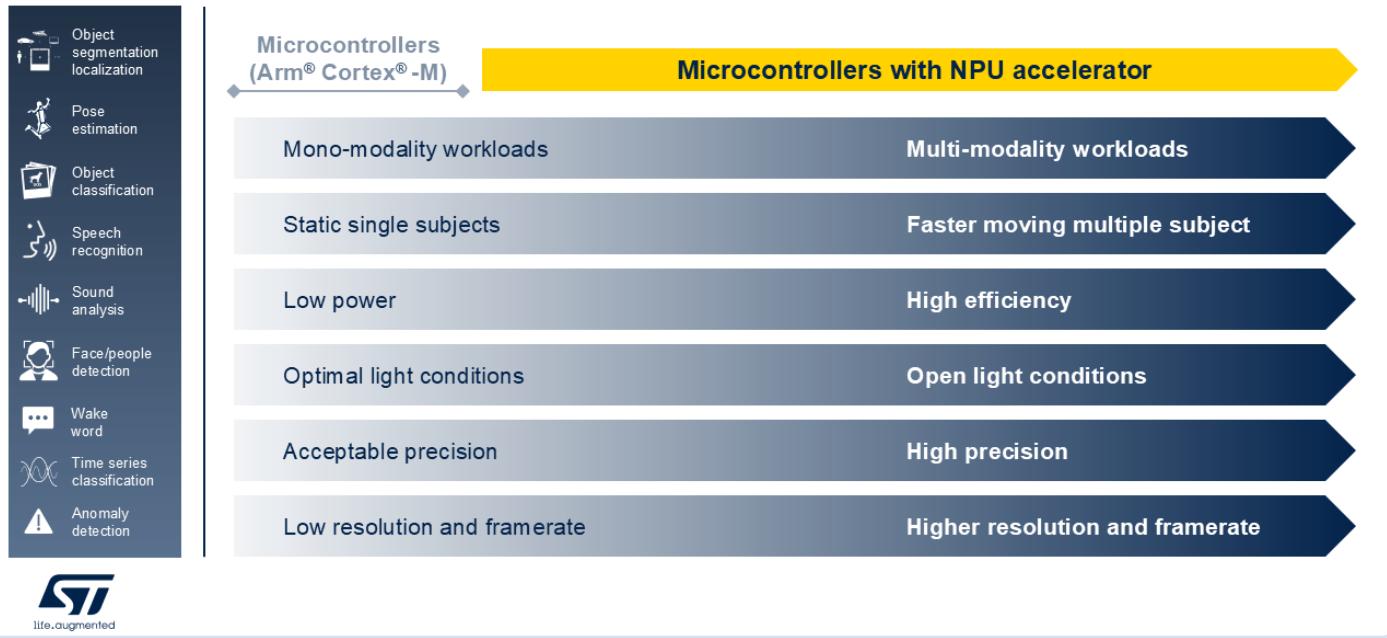


Figure 3. Le NPU加速器 offre une plus grande puissance de calcul pour les tâches d'intelligence artificielle, tout en maintenant une faible consommation d'énergie.

**Elektor :** Pouvez-vous nous en dire plus sur le ST Neural-ART accelerator ?

**François :** Le ST Neural-ART accelerator change la donne pour les applications d'IA en périphérie. Il s'agit d'un accélérateur matériel (NPU) conçu spécifiquement pour les réseaux neuronaux, qui offre un gain de performance de 600 fois supérieur environ par rapport à notre microcontrôleur STM32H7, tout en maintenant la consommation d'énergie à un niveau minimum (**figure 3**). Cet équilibre est essentiel pour les cas d'utilisation de l'IA dans les appareils alimentés par batterie, tels que les vêtements, les caméras intelligentes ou les capteurs industriels. Le NPU permet aux développeurs de déployer des modèles d'IA plus complexes, comme ceux nécessaires à la vision par ordinateur ou à la prise de décision en temps réel, directement sur des microcontrôleurs à faible consommation. Nous venons de publier un livre blanc pour en savoir plus à ce sujet et nous invitons les gens à le lire pour en savoir plus.

**Elektor :** Comment STMicroelectronics envisage-t-elle l'évolution de l'Edge AI ?

**François :** L'avenir de l'IA en périphérie est incroyablement prometteur. Nous allons voir plus d'appareils intelligents prendre des

décisions en temps réel, avec moins de dépendance à l'égard de l'infrastructure en cloud. Ce sera la clé pour construire des systèmes plus durables et plus réactifs, en particulier dans des domaines tels que les villes intelligentes, la santé et l'automatisation industrielle. Chez ST, nous nous concentrons sur la fourniture d'outils, de matériel et d'écosystèmes qui permettront aux développeurs d'innover et de repousser les limites de ce qui est possible avec l'IA en périphérie.

**Elektor :** Finalement, que pouvons-nous attendre de STMicroelectronics lors du prochain STM32 Summit ?

**François :** Le STM32 Summit [9], qui se tiendra le 10 décembre 2024, sera une formidable occasion pour les développeurs et les ingénieurs d'en savoir plus sur notre vision de l'Edge AI. Nous dévoilerons de nouveaux développements passionnants, notamment un aperçu de notre future feuille de route en matière d'IA, et de la manière dont nous soutenons davantage la communauté avec des outils et des ressources. J'invite tout le monde à se joindre à l'émission, pour en savoir plus sur notre stratégie Edge AI, et pour être aux premières loges pour voir comment ST stimule l'innovation dans ce domaine. 

VF : Laurent Rauber — 240621-04

## LIENS

- [1] Les microcontrôleurs STM32 : <https://www.st.com/stm32>
- [2] ST Edge AI suite : <https://www.st.com/st-edge-ai-suite>
- [3] STM32 Developer Zone : [https://www.st.com/content/st\\_com/en/stm32-mcu-developer-zone.html](https://www.st.com/content/st_com/en/stm32-mcu-developer-zone.html)
- [4] NEAI Studio : <https://stm32ai.st.com/nanoedge-ai/>
- [5] STM32 Cube.AI : <https://stm32ai.st.com/stm32-cube-ai/>
- [6] ST Edge AI Core : <https://www.st.com/en/development-tools/stedgeai-core.html>
- [7] ST Edge AI Developer Cloud : <https://stm32ai.st.com/st-edge-ai-developer-cloud/>
- [8] STM32 AI model zoo : <https://stm32ai.st.com/model-zoo/>
- [9] STM32 Summit : <https://www.st.com/stm32-summit>