



De Intel au Raspberry Pi

3 générations d'IA embarquée dans nos écrans DOOH

Gregoire Audibert – Co-fondateur & CEO, **Winter Mushroom**

Winter Mushroom, 1 av du champ de mars, 45100 Orléans



Winter Mushroom

QUI NOUS SOMMES

Des écrans qui adaptent leur contenu à chaque passant, en temps réel et de façon 100 % anonyme.

Plateforme DOOH

Affichage dynamique en points de vente, piloté en temps réel.

Présence internationale

France, Asie, Afrique francophone.

IA embarquée

Chaque écran adapte le contenu à son audience en fonction du genre et de l'âge

Pourquoi pas le cloud ?

LE PROBLÈME

Connectivité
Réseau instable en magasin

Latence
Délai incompatible avec le temps réel

Client perdu
Le passant est déjà parti

Conclusion
Inférence sur site nécessaire



Contraintes terrain

- Connectivité instable en magasin – le réseau ne suit pas.
- La latence réseau est incompatible avec le temps réel.
- Le passant est déjà parti quand la réponse revient.

❏ **Conclusion :** l'inférence doit se faire sur place, à l'intérieur de l'écran.

GÉNÉRATION 1

Intel Movidius NCS2



Plug & Play

Intégration immédiate via USB – aucune modification hardware nécessaire.



Coût faible

Accessible, facilement scalable pour un déploiement multi-écrans.



OpenVINO

Optimiseur de modèles Intel : conversion et compression pour la puce.



Comment fonctionne un optimiseur de modèles ?

Un optimiseur ne rend pas le modèle plus intelligent, il le rend plus efficace.

1

1. Convertir

Un format unique (graphe + poids) quel que soit le framework d'origine.

2

2. Simplifier

Retirer les opérations d'entraînement et précalculer ce qui est constant.

3

3. Fusionner

Regrouper les opérations enchaînées en une seule.

4

4. Réduire et compiler

Passer de 32 à 16 bits (voire 8) et compiler pour la puce (VPU du Movidius).

❏ Précision quasi identique, temps d'inférence divisé, exécutable sur une puce à quelques dizaines d'euros.

⚠ COUP DUR

Intel déprécie le Movidius NCS2

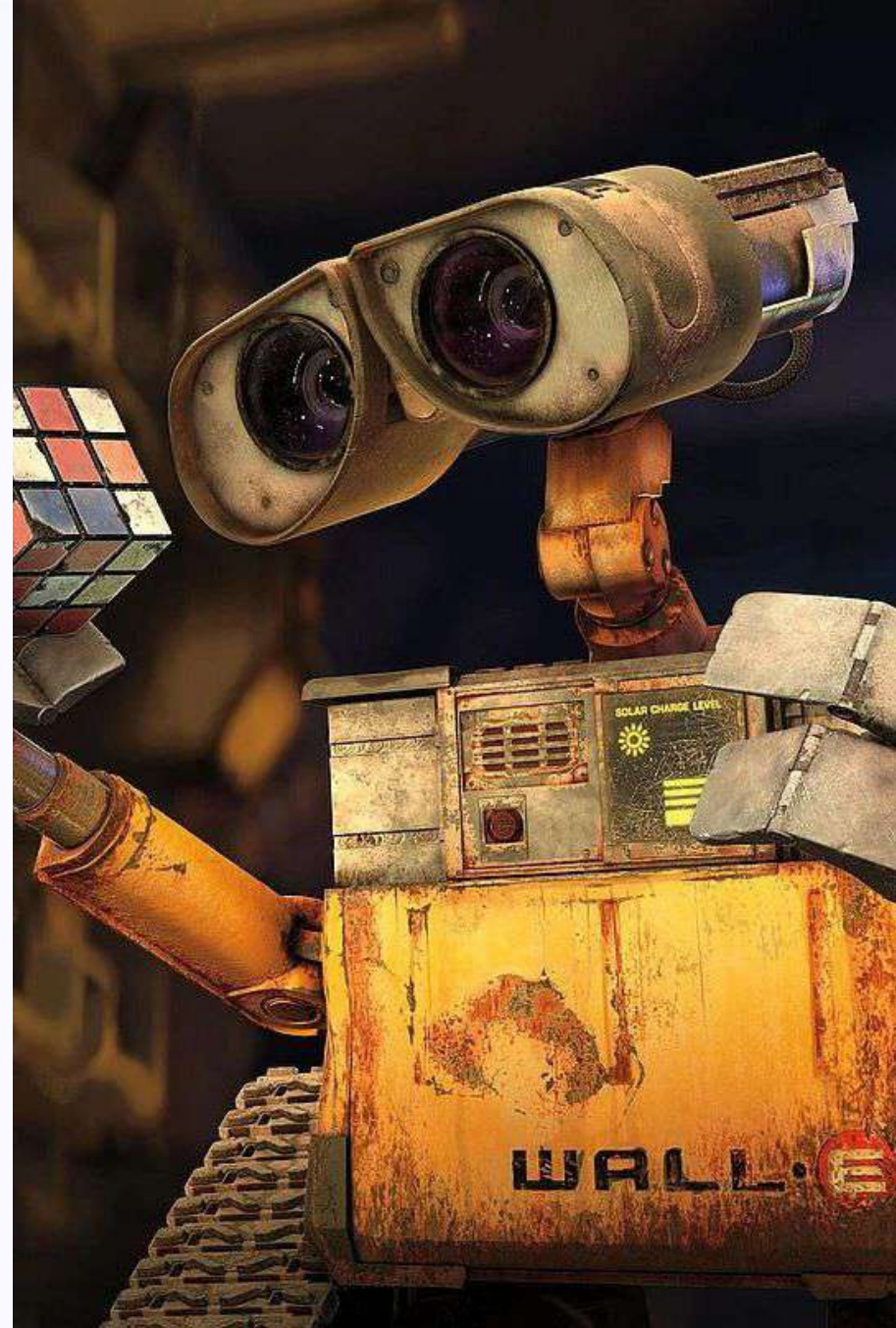
Du jour au lendemain, toute notre chaîne – matérielle *et* logicielle – doit être repensée. Rapidement.

Impact immédiat

Plus de support, le NCS2 coûte désormais environ 250€, la moitié du prix du media player.

Leçon retenue

Ne jamais dépendre d'un seul fournisseur de silicon.



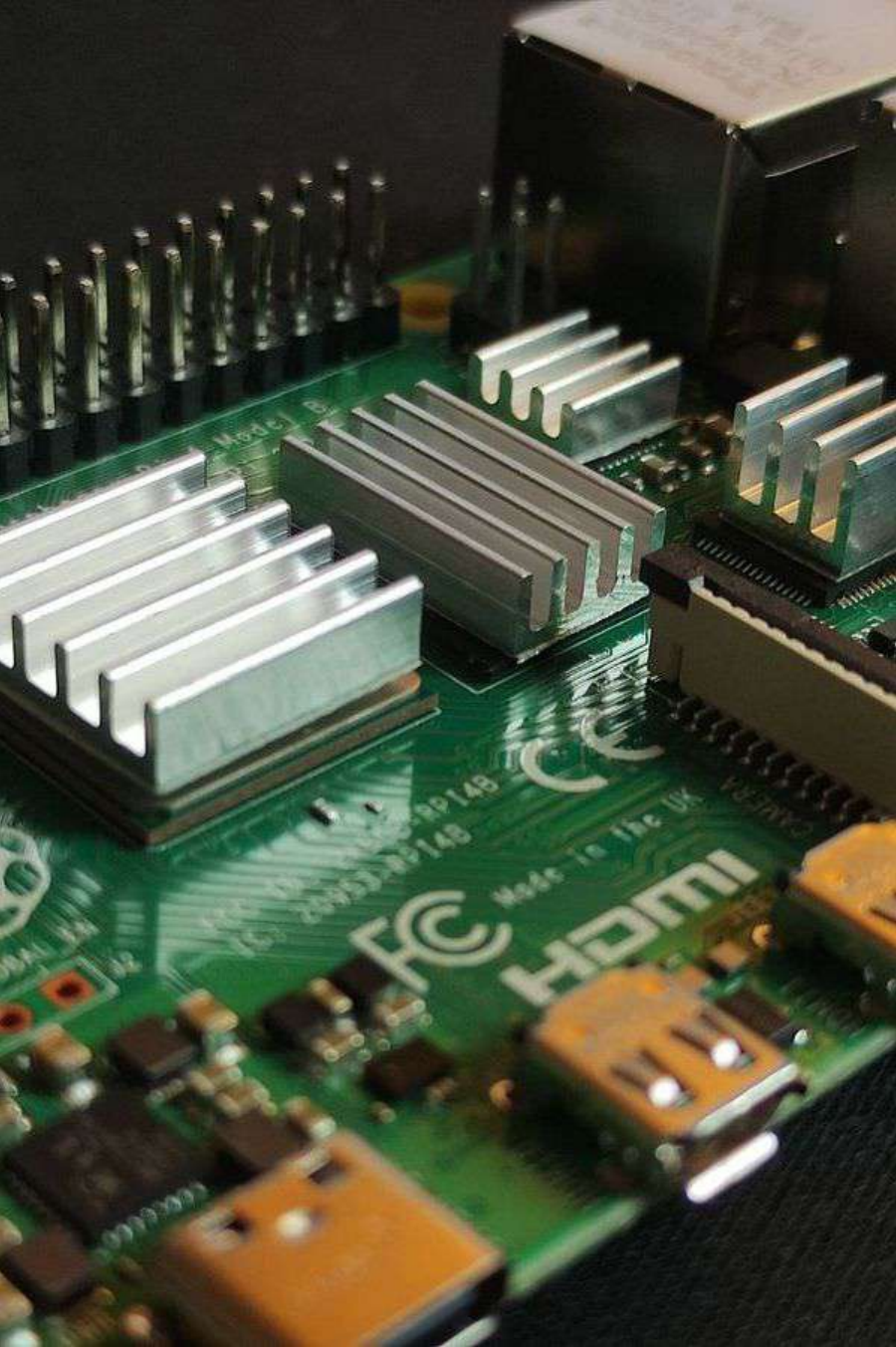


AU MÊME MOMENT

L'expansion internationale s'accélère

De nouveaux marchés s'ouvrent en Europe, Asie et Afrique francophone. Dans ces géographies, le **prix du matériel par écran** devient un critère stratégique d'entrée de marché.

- ❏ Un media player à 500€ est une barrière à l'entrée pour les pays émergents



GÉNÉRATION 2

Le pivot vers Raspberry Pi 5

Objectif

IA temps réel sur une machine abordable, disponible partout, sans accélérateur dédié.

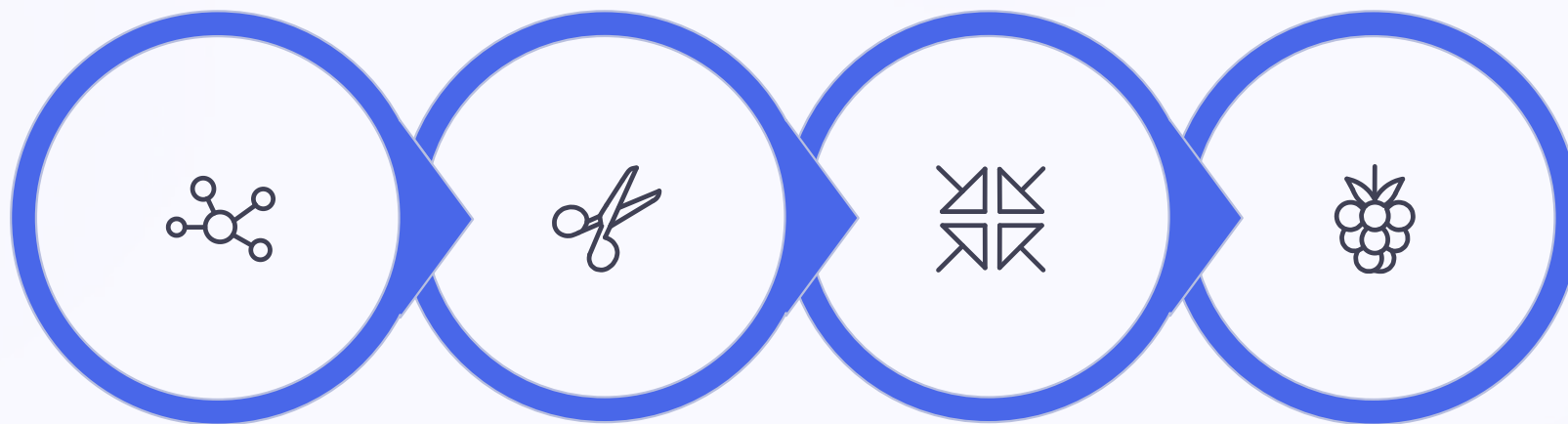
Le défi

Faire tourner en temps réel ce qui nécessitait auparavant une puce spécialisée.

Le résultat

Un modèle maison, entièrement repensé pour le CPU généraliste du RPi 5.

Apprendre grand, déployer petit.



Modèle 2×

Élagage 50%

Quantification 8 bits

Déploiement Pi 5

Entraîner grand – modèle 2× surdimensionné, puis élagage (~50 % des poids) et quantification 32 → 8 bits.

Accélérateur dédié ou optimisation logicielle ?

Nous sommes un éditeur logiciel. Notre valeur doit vivre dans notre code, pas dans une puce que nous ne maîtrisons pas.

Accélérateur dédié

- Dépendance au SDK et formats propriétaires du fabricant
- Feuille de route subie (vécu avec le NCS2)
- Coût supplémentaire sur chaque écran, multiplié par tout le parc

Optimisation logicielle

- Modèle sur CPU standard (Raspberry Pi aujourd'hui, autre demain)
- Mises à jour déployées à distance sur tout le parc
- Savoir-faire qui s'accumule d'une génération à l'autre

- ❏ Le compromis assumé : un accélérateur irait plus vite, mais un CPU bien exploité suffit pour détecter âge et genre en temps réel.



Merci – Des questions ?

Gregoire Audibert

Co-fondateur & CEO

Winter Mushroom

Plateforme DOOH avec IA embarquée