

Les 12 règles d'or pour des vidéos IA cohérentes

Tirées de six épisodes de SerialKiwix : le pourquoi, l'exemple vécu et le prompt ou le réglage qui marche

Par l'équipe de SerialKiwix, la série
fabriquée avec l'IA.
fruisse.com/ecole · octobre 2026



Une vidéo IA ratée, c'est rarement une image moche. C'est un personnage qui change de visage entre deux plans, un objet qui passe du sol à la main sans que personne l'ait ramassé, une réaction à un bruit qu'on n'a pas entendu. Le spectateur ne saurait pas toujours dire ce qui cloche ; il sait qu'il décroche.

Ces douze règles viennent de notre production réelle : **SerialKiwix**, la série où Kiwix, barman du bar à jus « Pressé ! », presse la nuit les fruits qui l'ont mérité (sans une goutte de sang : du jus). Après chaque épisode, on écrit un retour d'expérience, et chaque erreur devient une règle avec une lettre. Six épisodes plus tard, on en compte 35, de A à Al. On a gardé ici les douze qui changent le plus la cohérence, rendues valables pour n'importe quel projet : série, pub, clip ou histoire courte.

Pour chacune : **pourquoi** elle existe, **ce qui nous est arrivé**, et **le prompt ou le réglage** à reprendre tel quel. Les lettres entre crochets renvoient à nos règles d'origine.

12 RÈGLES D'OR DE COHÉRENCE

Toutes nées de nos erreurs sur SerialKiwix (épisodes 1 à 5).
À relire **avant** chaque génération.

LES PERSONNAGES ET LE DÉCOR

- 1** Un bloc canonique par personnage, copié mot pour mot.

Kiwix : « **golden-brown fur, not olive, no beard** ».

- 2** Plan large d'abord, gros plans à partir de lui.

La salle de bain de Pêche sert de référence à **tous** ses plans.

L'HISTOIRE TIENT DEBOUT

- 3** Écris la chaîne « cause → effet » avant de générer.

L'écharpe va à la fenêtre → Kiwix grimpe → **mixeur ON** → Pêche crie.

- 4** Toute réaction a sa cause à l'écran.

Mangue **entend le cri** (un plan dédié), puis elle arrive.

- 5** Qui ne doit pas voir le tueur ne le touche pas.

vl : Pêche versait du shampoing sur Kiwix... **sans être censée le voir**.

- 6** Note l'état du décor après chaque événement.

Rideau tombé, fenêtre ouverte : **repris** dans tous les plans suivants.

- 7** Suis chaque objet clé d'un plan à l'autre.

vl : le mixeur posé au sol... **puis tenu en l'air**.

- 8** L'accroche reprend les plans de sa scène.

vl : un mixeur de la **cuisine de l'E04** dans l'accroche de l'E05.

LES IMAGES ET LE CONTRÔLE

- 9** Aucun texte généré par l'IA.

« **blank paper, no writing** » : le texte s'ajoute au montage.

- 10** Recalcule les comptes à rebours.

« Plus que sept jours » répété le lendemain → « **Plus que six jours** ».

- 11** 2 graines pour chaque plan d'action important.

Kiwix **revenait dans la fenêtre** au lieu de fuir.

- 12** Relis chaque vidéo sur 5 images, du début à la fin.

0,2 · 0,8 · 1,6 · 2,6 · 3,6 s : le rideau se dessinait **des yeux à la fin**.

Le test final : regarde l'épisode en entier une fois en te demandant « pourquoi ? » à chaque plan.

Source : règles A à AI des retours d'expérience SerialKiwix, checklist de cohérence.

fruisse.com

Règle 1 – La chaîne logique s'écrit avant la première image [AB]

Pourquoi. Les incohérences les plus coûteuses ne sont pas des ratés de génération, ce sont des trous dans l'histoire. Un modèle ne sait pas que ton personnage devait entrer par la fenêtre ; il dessine ce que tu lui demandes, plan par plan. Si l'enchaînement n'existe que dans ta tête, il n'existe pas à l'écran.

Chez SerialKiwix. L'épisode 5, « La douche », a été fabriqué trois fois dans la même journée. Environ **40 % du temps machine** a servi à corriger des erreurs de logique : on ne voyait jamais Kiwix entrer, la lieutenant Mangué arrivait sans avoir entendu le cri, le hook utilisait un plan de mixeur pris dans un autre épisode. Toutes auraient été vues sur le papier.

Le réglage. Pour chaque scène, une ligne « cause → effet », avant de générer quoi que ce soit. La nôtre, telle qu'elle a été validée :

```
L'écharpe va du panier à la fenêtre → Kiwix grimpe et entre par la fenêtre  
→ mixeur ON → l'écharpe tombe dedans et le casse → Pêche crie, le rideau  
tombe sur elle → Mangué entend → Kiwix fuit par l'écharpe → Mangué entre :  
pièce vide.
```

Ajoute : par où chacun entre et sort, où est chaque objet, qui voit quoi, et **quels plans** le hook reprend (uniquement ceux de la scène qu'il annonce [AC]).

Règle 2 – Un bloc canonique en anglais, recopié mot pour mot [P]

Pourquoi. L'IA n'a pas de mémoire. Chaque prompt est une nouvelle demande ; trois façons de décrire ton personnage donnent trois personnages. La seule mémoire possible, c'est un texte figé, recopié à l'identique.

Chez SerialKiwix. Kiwix a le même paragraphe d'apparence dans chaque prompt, et ses tenues dans des blocs séparés. On l'a même appliqué à la miniature, l'image qu'on fabrique en dernier et que tout le monde voit en premier.

Le prompt.

an anthropomorphic kiwi fruit: egg-shaped body covered in scruffy matted golden-brown kiwi fur, large bulging round eyes with bright green irises, heavy furrowed brows, thick black handlebar mustache, wide unsettling toothy grin, thin furry brown arms, stubby furry legs; warm golden-brown fur, not olive, no beard

JOUR : lime-green apron over a white short-sleeved shirt, small dark green bow tie
NUIT : worn torn beige trench coat, black fedora hat, white latex gloves

ANATOMIE D'UN BLOC CANONIQUE

Une description en anglais, copiée **mot pour mot** dans chaque prompt : c'est elle qui garde Kiwix identique d'un plan à l'autre.



- 1 CORPS**
La forme exacte : le fruit entier, petits bras, petites jambes. Jamais un corps humain.
- 2 YEUX**
Taille, forme, couleur : ce que l'IA change en premier.
- 3 SIGNES DISTINCTIFS**
Ce qui le rend reconnaissable en une seconde.
- 4 COULEURS**
Écrites noir sur blanc, nuance comprise.
- 5 TENUE**
Copiée elle aussi, mot pour mot.
- 6 INTERDITS**
Ce que l'IA ajoute toute seule si on ne dit rien.

LE VRAI BLOC DE KIWIX

tenue de nuit

an anthropomorphic kiwi fruit: egg-shaped body covered in scruffy matted golden-brown kiwi fur ①, large bulging round eyes with bright green irises, heavy furrowed brows ②, thick black handlebar mustache, wide unsettling toothy grin ③, thin furry brown arms, stubby furry legs ①; warm golden-brown fur ④, not olive, no beard ⑥, wearing a worn torn beige trench coat, black fedora hat, white latex gloves ⑤

Tenue de jour, au bar « Pressé ! » : lime-green apron over a white short-sleeved shirt, small dark green bow tie

A Copie le bloc **mot pour mot** dans chaque prompt, tenue comprise. Ne le réécrit jamais « à peu près ».

B Ajoute 2-3 images de référence et « **exact same face, colors and eye size as <image2>** ».

C Pas d'adjectif extrême seul (« **creepy** ») : décris l'expression, comme « **wide unsettling toothy grin** ».

Source : fiche de Kiwix et prompts réels des épisodes SerialKiwix.

fruisse.com

L'ordre : la nature, la silhouette et la matière, le visage du plus voyant au plus fin, les membres, puis les **garde-fous**. Chaque défaut qui revient deux fois y gagne son interdit : « not olive » et « no beard » sont nés de vraies dérives.

💡 Astuce : range tes blocs dans un fichier et copie-colle, ne retape jamais. Un mot changé (« big » au lieu de « bulging »), c'est un visage changé.

Règle 3 – Des traits visibles, jamais des adjectifs d'humeur [B]

Pourquoi. « Creepy », « evil », « psychopath » : le modèle traduit l'humeur en exagération, et chaque image exagère à sa façon. Ton personnage devient un monstre différent à chaque plan.

Chez SerialKiwix. Kiwix doit inquiéter tout en restant le même barman poli. Ce qui marche : décrire **ce qu'on voit** sur son visage, pas ce qu'on doit ressentir.

Le prompt.

```
wide unsettling toothy grin  
polite smile that is slightly too wide
```

Vaut aussi pour l'action : « comedic food-pressing », « food commercial style », « cartoon, no blood » plutôt que des mots comme kill, murder, blood ou explosion, qui font dériver l'image et déclenchent les filtres des services en ligne.

Règle 4 – Des références propres, ordonnées, jamais remplacées [B, I]

Pourquoi. L'image de référence ancre le visage ; le texte ancre les détails. Mais une référence transmet **tout**, y compris ses défauts.

Chez SerialKiwix. Kiwix est généré avec deux ou trois références (planches officielles et plans validés tirés des épisodes). Une de ses références au bar recopiait l'inscription « FRESH SQUEEZE » dans les images suivantes : elle est barrée dans notre registre, « NE PLUS UTILISER ». Et sur une image à trois références, la robe crème de Pêche a disparu alors qu'elle était sur sa référence : ce qui compte s'écrit aussi.

Le réglage. Toujours le même ordre : `<image1>` le décor (ou le plan large), `<image2>` le personnage, `<image3>` le second personnage ou l'objet. Jusqu'à quatre références sans souci chez nous. Et cette phrase :

```
exact same face, fur color and eye size as the reference
```

△ Piège : changer de référence en cours de saison « parce qu'une autre est plus jolie ».
Tout ce qui a déjà été fabriqué ne lui ressemblera plus.

Règle 5 – Le plan large d'abord, les gros plans à partir de lui [B, U]

Pourquoi. Un gros plan généré de zéro réinvente la lumière, le fond, les accessoires. Généré à partir du plan large, il en hérite.

Chez SerialKiwix. C'est l'une des toutes premières règles, née du pilote, et la plus constante. Le décor maître de la salle de bain de l'épisode 5 a été fabriqué à part, avant les plans, puis donné en première référence de chacun d'eux.

PLAN LARGE D'ABORD, GROS PLANS ENSUITE

La méthode qui garde le même décor dans tous les plans d'une scène. Exemple réel : la salle de bain de Pêche, épisode 5.



1 LE DÉCOR MAÎTRE

Le lieu, généré **une seule fois**. Il devient la référence **<image1>** de tous les plans de la scène.



2 LE PLAN LARGE DE LA SCÈNE

Toute l'action : qui est où, quels objets, quelle lumière. Ici, l'écharpe rose part du panier vers la fenêtre.

PROMPT
Same bathroom as <image1>. Wide shot...



3 LES GROS PLANS, À PARTIR DU PLAN LARGE



Le rideau opaque : une silhouette floue, rien d'autre.



Kiwi et le mixeur, dans la même pièce.

Référence = le plan large + l'image du personnage.

PROMPT
TIGHT close-up, face and shoulders only, filling the frame

CE QUI RESTE IDENTIQUE D'UN PLAN À L'AUTRE

✓ la lumière

✓ le fond

✓ les accessoires

✓ l'état du décor

✗ **Sans cette méthode**, chaque gros plan réinvente son décor : autre lumière, autre fond, le panier et l'écharpe disparaissent.

Astuce : jusqu'à 4 références par image. Le décor d'abord (<image1>), les personnages ensuite (<image2>, <image3>...).

Source : guide de production locale et images réelles de l'E05.

fruisse.com

Le prompt. Le plan large en <image1>, le personnage en <image2> :

Same scene as <image1>. TIGHT close-up, face and shoulders only, filling the frame. The kiwi barman from <image2> (<bloc>; <tenu>) ... MOUTH CLOSED, eyes open. Soft golden morning light from the window, warm golden-brown fur, not olive.

La lumière s'écrit, couleur du personnage comprise [U] : un éclairage peut repeindre un personnage.

Règle 6 – Chaque personnage a sa place et sa taille dans le cadre [L, Q, S, X]

Pourquoi. Un personnage trop petit n'est pas reconnaissable ; un personnage mal placé contredit le plan suivant ; dans un plan à plusieurs, celui qu'on ne situe pas, le modèle le place où il veut.

Chez SerialKiwix. Nos règles : personnages au moins au **tiers de la hauteur** dans les plans larges, premier plan serré sur le sujet [Q], **quatre personnages au maximum** par plan large, place de chacun écrite [L], un **plan à deux** avant les champs-contrechamps [S], et le personnage indispensable vérifié dans ce plan à deux avant de dériver les gros plans [X]. Sur l'épisode 6, Kiwix est sorti debout **sur** son comptoir jusqu'à ce qu'on écrive où sont ses pieds.

Le prompt.

clearly visible, filling 40 % of the frame height
stands BEHIND the counter / customer side, in front of the counter
standing on the FLOOR BEHIND the counter... legs hidden
Foreground, right side, medium close shot: ... Background, deep in the
alley, smaller but clearly readable: ...

Règle 7 – L'objet clé est nommé et en référence dans chaque plan [K, V]

Pourquoi. Un objet qui n'est qu'un mot est réinventé à chaque image. Et un objet qui change de forme au milieu d'une scène casse la continuité aussi sûrement qu'un visage.

Chez SerialKiwix. Les armes de Kiwix sont des ustensiles de cuisine, et chacune est nommée et donnée en référence dans tous les plans de sa scène. Sur la miniature de l'épisode 5, sans référence, le modèle a dessiné un **épluche-légumes** à la place du mixeur.

Le prompt. L'objet en <image3>, sa forme décrite, et pour les objets à risque :

Règle 8 – L'état du décor (et du personnage) après chaque événement [AF, H]

Pourquoi. Ce qui est cassé reste cassé, ce qui est tombé reste tombé. Le modèle, lui, remet tout en ordre à chaque image si tu ne lui dis rien.

Chez SerialKiwix. Dans la v1 de l'épisode 5, le rideau était intact et la fenêtre fermée quand Mangue entrait, juste après la catastrophe. Depuis, la fiche de continuité de chaque scène se termine par **l'état du décor après**, repris dans tous les plans suivants et le lendemain : le mixeur cassé reste dans la salle de bain, et la police l'y retrouve. Même chose pour le personnage : Kiwix repart avec de la laine rose dans la fourrure, et un brin est encore sur son épaule au bar le lendemain. Et l'échelle se respecte aussi après coup [H] : les restes d'une victime sont à sa taille.

Le réglage. Une ligne à la fin de chaque fiche de scène, copiée dans les prompts suivants :

État du décor APRÈS : rideau tombé, fenêtre ouverte, mixeur cassé au sol

Règle 9 – Qui voit quoi : chaque regard et chaque réaction ont leur cause [J, AD, AE, N]

Pourquoi. Un regard vers rien, une réaction sans cause ou un personnage qui interagit avec quelqu'un qu'il n'est pas censé voir : le spectateur le sent immédiatement, même s'il ne sait pas le nommer.

Chez SerialKiwix. Ce qu'un personnage regarde est visible [J]. Toute réaction a sa cause à l'écran [AD] : Mangue entend le cri, on la voit entendre, puis elle arrive. Pêche ne doit pas voir Kiwix, donc elle n'a aucune interaction directe avec lui [AE] (dans la v1, elle lui vidait son shampoing dessus). Et une réaction est une **image distincte**, sous un autre angle [N].

Le prompt. Le regard s'écrit, comme une action :

staring into the alley without blinking
He has not noticed her.

Règle 10 – Zéro texte généré par l'IA [E, T]

Pourquoi. Les modèles écrivent mal : lettres inventées, mots à moitié faux, et parfois du texte qui n'a rien à faire là. Un panneau illisible ou en mauvaise langue suffit à briser l'illusion.

Chez SerialKiwix. Notre enseigne « Pressé ! » est sortie avec un « ! » parasite au début ; un panneau « Closed » en anglais s'est invité dans un décor français (épisode 6). Depuis le pilote, tout texte à l'écran est composé à part, lisible et gratuit : la une du journal « Le Petit Pressé », les contrats, les comptes à rebours, l'habillage de la télé de Cerise. Et la rubalise de la police est « printed only with BCM ».

Le prompt.

```
blank paper, no writing  
no text, no banner, no overlay  
No text, no letters, no logo, no watermark.
```

Règle 11 – Le son colle à l'image [C, AG, AH]

Pourquoi. Une bouche qui ne suit pas la voix, un « pop » qui tombe à côté de la bulle, une ambiance qui siffle : le son casse la cohérence aussi sûrement qu'une image.

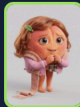
Chez SerialKiwix. Les plans parlants partent d'une image bouche fermée, caméra fixe [C]. En local, la réplique est rallongée de silence jusqu'à **5,6 secondes**, sinon le modèle invente une suite en anglais. Le « pop » de la bulle de Mangue était placé à 2,0 s alors que la bulle éclate à 2,25 s : un quart de seconde, et le gag tombait à plat [AG]. Notre première douche était stridente ; refaite douce et coupée vers 2,8 kHz, elle passe [AH].

Le réglage.

```
Image de départ : MOUTH CLOSED, eyes open, static camera  
Prompt vidéo : <personnage> is speaking with natural lip movements  
synchronized, then stays silent with the mouth closed.  
Réplique : rallongée de silence à 5,6 s ; jamais les mots dans le prompt  
Eau, pluie, vapeur : passe-bas vers 3 kHz ; aucun son sans image  
Mixage : ≈ -14 LUFS, pic sous -0,3 dB, audio en 48 kHz
```

LA SYNCHRO LABIALE EN LOCAL

Faire parler un personnage avec MiniMax H3 sur ton PC, sans qu'il invente la suite. Toutes les pistes sont à la même échelle.

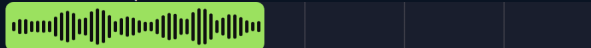


PÊCHE, ÉPISODE 5

« J'ai rien vu... mais on a mixé mon tricot ! »

0 s 1 s 2 s 3 s 4 s 5 s 6 s

1 LA RÉPLIQUE ELEVENLABS



2 RALLONGÉE DE SILENCE JUSQU'À 5,6 s



3 MINIMAX H3 REÇOIT TROIS CHOSES

IMAGE DE DÉPART

« MOUTH CLOSED,
eyes open, static
camera »

AUDIO_GUIDE

la réplique
rallongée à 5,6 s

LE PROMPT

il décrit le jeu, pas le
texte

<personnage> is speaking with natural lip movements
synchronized, then stays silent with the mouth closed.

Ne jamais écrire les mots de la réplique.

4 LA VIDÉO DURE ≈ 4,5 s

107 images minimum : H3 ne fait jamais plus court.



5 COUPE AU MONTAGE

On garde jusqu'à la fin de la réplique : après la voix, le personnage change d'expression.



× fin de la réplique



Sans la rallonge de silence, H3 invente une suite à la réplique... souvent en anglais.

Source : guide de production locale et retour d'expérience de l'E05.

fruisse.com

Règle 12 – Relis en planches, avec deux graines pour les plans qui comptent [G, AI]

Pourquoi. Un défaut de cohérence ne se voit pas sur une image isolée : il se voit **à côté** de la précédente. Et un plan vidéo peut commencer juste et dériver à la fin.

Chez SerialKiwix. Images relues en planche contact, côte à côte ; vidéos relues sur **cinq images** (0,2 / 0,8 / 1,6 / 2,6 / 3,6 s) ; montage relu en planches d'une image par seconde [G]. C'est ainsi qu'on a attrapé Kiwix qui revenait dans la fenêtre qu'il venait de fuir, et un rideau uni qui se dessinait un visage. Pour les plans d'action importants, **deux graines d'emblée** [AI] : en local, une minute et demie. Sur l'épisode 6, la relecture complète a encore trouvé un rideau de fer ouvert après sa fermeture : l'ordre des plans était inversé.

Le réglage. Pour extraire les cinq images d'un plan :

```
for t in 0.2 0.8 1.6 2.6 3.6; do
  ffmpeg -loglevel error -ss $t -i P05.mp4 -frames:v 1 P05_$t.png
done
```

Puis un visionnage complet, en se demandant « **pourquoi ?** » à chaque plan.

Les 12 règles sur une page

#	Règle	La formule à retenir
1	Chaîne logique avant la première image	« cause → effet » par scène ; hook = plans de la scène
2	Bloc canonique mot pour mot	nature → silhouette → visage → membres → garde-fous
3	Traits visibles, pas d'humeur	« polite smile that is slightly too wide »
4	Références propres et ordonnées	« exact same face, fur color and eye size as the reference »
5	Plan large d'abord	« TIGHT close-up, face and shoulders only, filling the frame »
6	Place et taille écrites	« clearly visible, filling 40 % of the frame height »
7	Objet clé en référence	« holds NOTHING else »
8	État du décor après	« rideau tombé, fenêtre ouverte... » repris partout
9	Qui voit quoi	« He has not noticed her. »
10	Zéro texte généré	« blank paper, no writing »
11	Le son colle à l'image	bouche fermée, réplique à 5,6 s, bruitage à l'instant exact
12	Planches et deux graines	cinq images par vidéo, « pourquoi ? » à chaque plan

✓ À retenir : la cohérence ne vient pas d'un meilleur modèle, elle vient de ce que tu lui donnes, à l'identique, et de ce que tu relis côte à côte. Douze règles, et chacune tient en une phrase à coller dans ton prompt ou dans ta checklist.

Les outils cités, et leurs licences

Nos réglages viennent de notre chaîne locale : Qwen Image 2.1 pour les images, MiniMax H3 avec le LoRA « Pika merge lite » pour les vidéos, sur une RTX 5080. Les règles, elles, valent pour n'importe quel outil, local ou cloud.

⚠ Piège : les licences (lues le 31/10/2026). **Qwen Image 2.1** est sous licence de recherche, **non commerciale** ; la licence **MiniMax H3 Community** (qui couvre aussi Pika H3 Flex et le merge lite) **exclut de son territoire l'Union européenne**, le Royaume-Uni, la Corée du Sud et les États-Unis. Alternatives : **LTX-2.5** (usage commercial gratuit sous 10 M\$ de chiffre d'affaires annuel, pas encore testé par nous) ou les services cloud, dont les conditions commerciales sont à lire chez chacun. Lis la licence de l'outil que tu choisis avant d'y construire ta série.

Pour aller plus loin

Ce guide te donne les règles. Le cours « **Un personnage qui reste le même, plan après plan** » de l'École Fruise te donne la méthode complète pour les appliquer : la fiche personnage détaillée, la construction pas à pas d'un bloc canonique (avec nos blocs disséqués), le choix et l'entretien des références, la fabrication d'une scène entière du décor maître aux gros plans, les 35 règles de A à AI expliquées, un catalogue de quinze ratés réels avec leur correction, et la checklist de cohérence à remplir avant chaque lot. La première leçon est en accès libre ; les suivantes s'ouvrent avec le Pass École.

Et pour la chaîne complète, de l'idée à la publication, il y a le guide gratuit et le cours « **Ta première vidéo IA en 7 étapes** ».

Aucun pépin.

Va jusqu'au bout avec le Pass École

Ce guide te donne la carte. Les cours de l'École Fruise te donnent le pas-à-pas : nos vrais prompts et réglages, les infographies, les vidéos expliquées, les budgets et les projections chiffrées.

29 €

Tous les cours pendant 30 jours. Paiement unique, sans renouvellement automatique.

Commence par les leçons gratuites : fruisse.com/ecole

Le Pass : fruisse.com/ecole/pass