Engraved by Program: A Study of Language Models Writing and Revising Procedural Drawings of Plant Species

Jason Gardner

Independent researcher · https://electrohead.net · ORCID 0009-0005-3882-0025

Working paper, September 2026. Not peer reviewed.

Abstract. Can a language model look at reference images of a plant, write what it sees as procedural drawing rules, and revise the rules until the drawing is recognisable as that species? We study this on a recorded corpus: all 33 plants drawn for poisongarden.org, the companion site to a book in progress on naturally occurring poisons and venoms, each a seeded program that emits engraving-style vector strokes, 16,276 lines in all, written entirely by language models under a render-and-revise loop with a human owner. The corpus was made in three phases that differ in what the model had in hand, from botanical knowledge alone to plates and eye-level photographs, and every version and most rounds of revision survive. Five case studies follow the model from references to program, with the mathematics of each construction it tried and each that replaced it: a one-sided raceme, recursive growth under a crown envelope, a groove reserved on a dark ground, a gaussian-notched leaf, and a projected frond. Across the corpus we coded 210 drawing defects with two blind coders ($\kappa = 0.88$ for kind). Programs written from knowledge alone scored lower in a review against plates (median 3 against 4 of 5), and only they broke the width budget. Most defects were found by sight rather than measurement, and among the fixed defects one found by sight had 5.4 times the odds of a structural fix, a replaced construction rather than a tuned constant. Where the model's first guesses failed, they were textbook forms, true to how a plant is built but not to how it looks.

Keywords: procedural modelling, botanical illustration, large language models, program synthesis, iterative refinement, failure analysis, non-photorealistic rendering.

1 Introduction

A botanical illustration is judged by one question: does the reader recognise the plant? Procedural modelling has generated plausible plants from rules for half a century [1, 2, 3], and models of particular species, calibrated from measurements [4] or rebuilt from photographs [5, 6], are well established. Their rules are mostly written by people who know both botany and programming, and they aim at accurate structure for science or rendering. Large language models now write graphics programs [7, 8, 9], can look at the images those programs produce [10, 11], and have begun to help write plant models [12, 13]. We study one use of that combination, aimed at illustration rather than reconstruction: *can a language model look at reference images of a particular plant, write the patterns it sees as procedural drawing rules, and revise the rules, render by render, until the drawing is recognisable as that species?* And if it can, where does it go wrong, who notices, and what does the fix look like?

We study this question on a real corpus rather than a benchmark: all 33 plants drawn for poisongarden.org, each a seeded program, 16,276 lines of code in all, every line written by language models over two weeks under a render-and-revise loop with a human owner (Fig. 1). The corpus has two properties that make it useful for research. It was written in three phases that differ in what the model had in hand, from botanical knowledge alone to plates and eye-level photographs, which gives a natural comparison. And its making was recorded: every version of every program, and for most plants the defects named in each round and the transcripts of the sessions that named them.

We ask four questions.

RQ1 Does drawing from reference images, rather than from knowledge, change how recognisable the first program is?

RQ2 What goes wrong when a language model writes a procedural drawing of a species, and are the failures found by measurement or by sight, and fixed by tuning or by rebuilding?

RQ3 Who finds the defects that matter: the model that wrote the code, a model that reviews it, or the owner?

RQ4 What constructions carry recognisability, and how do they differ from the model's first guesses?

To answer them we traced five plants through their full histories, from the references to the final program, with the mathematics of each construction that was tried and each that replaced it (Section 4), and we coded all 221 defects named across the 33 plants, 210 of them defects in the drawing itself (Section 5). The five were chosen to span what the model had in hand and how the plant is built: foxglove, drawn from knowledge alone; the Jerusalem cherry, rejected when drawn from text and rebuilt from references; the Calabar bean, for which no photographs exist; bloodroot, whose photographs

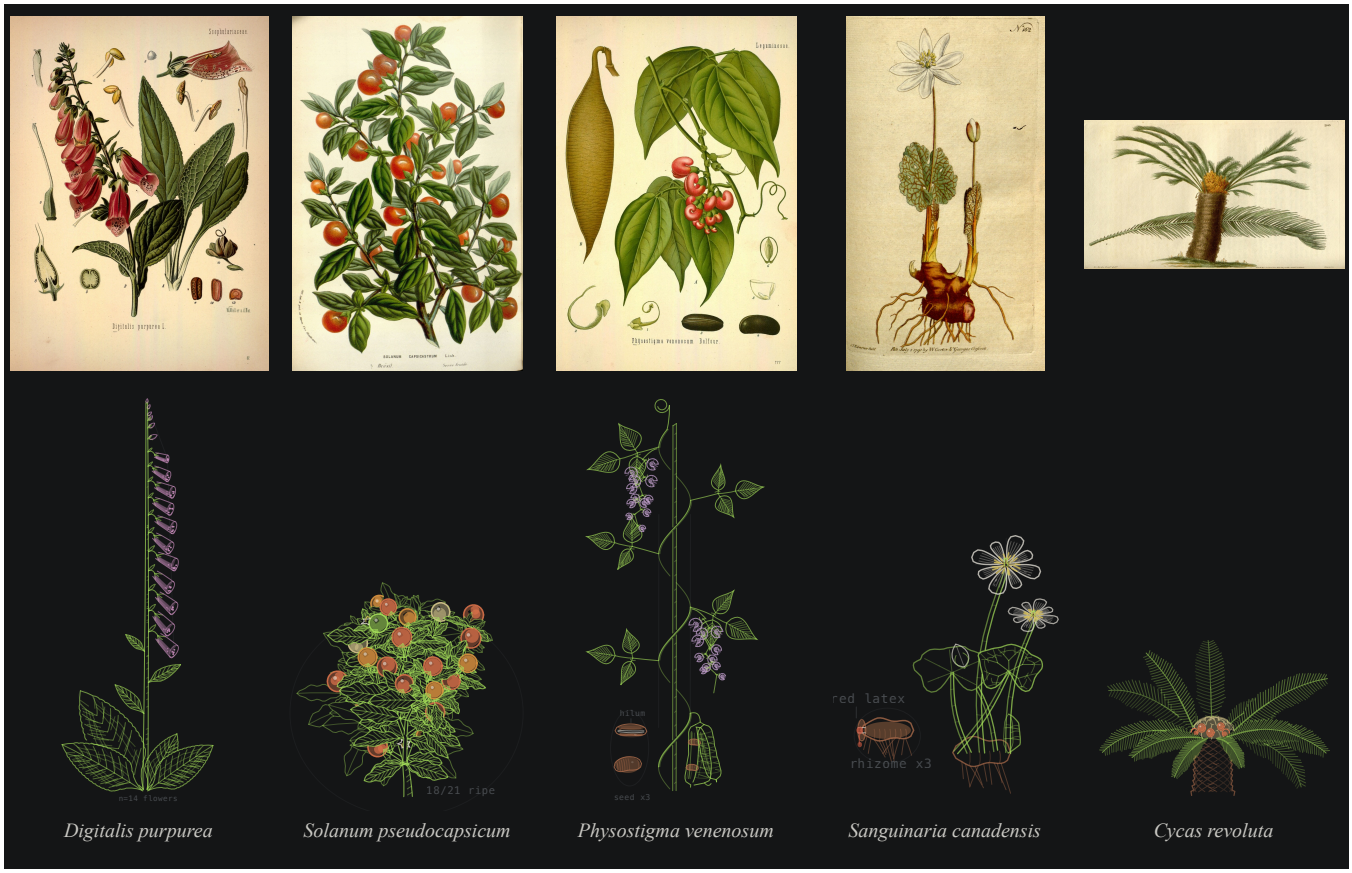


Figure 1: The five case-study plants. Top: the historical plates each drawing was checked against. Bottom: the final drawings, each the output of its own procedural program at seed 11 and detail 0.9, each scaled to fit its column. From left: foxglove (*Digitalis purpurea*; Köhler's *Medizinal-Pflanzen*, 1887), Jerusalem cherry (*Solanum pseudocapsicum*; Van Houtte's *Flore des serres*, vol. 12, 1857, plate 1242, under the synonym *S. capsicastrum*), Calabar bean (*Physostigma venenosum*; Köhler, vol. 2, 1890, whose coral-red flowers are an error of the plate; the drawing follows the written descriptions), bloodroot (*Sanguinaria canadensis*; Sydenham Edwards, *The Botanical Magazine*, 1792) and sago palm (*Cycas revoluta*; *Curtis's Botanical Magazine*, plate 2963). All plates are in the public domain.

were taken from the wrong angle; and the sago palm, drawn last, under the settled protocol.

Our findings, in brief. Plants drawn from knowledge alone scored lower in a separate model's review against plates, and only they broke the width budget. The model's geometry was rarely the problem; its first guess at an organ was the textbook form, a cardioid leaf, a phyllotactic spiral of flowers, radial veins under rings, which describes how the plant is built but not how it looks. Most defects were found by sight rather than measurement, and those that were fixed were mostly fixed by replacing a construction rather than tuning it. The builder named half the defects itself, but the ones that sent a drawing back were named by a critic that had not written it.

The contributions of this paper are:

1. A study of language models writing and revising procedural drawings of real species, on a recorded corpus of 33 programs, 61 versions and 221 coded defects, with a codebook and reliability measures (Sections 3 and 5).
2. Five case studies that follow the model's reasoning from references to program, with the mathematics of first and final constructions: view geometry, recursive growth under a crown envelope, tone on a dark ground, a reserved groove, a gaussian notch, a one-sided raceme and a projected frond (Section 4).
3. Evidence on grounding, failure and critique (RQ1–RQ3) across the whole corpus (Section 5).

2 Related Work

2.1 Procedural modelling of plants

Procedural plant modelling began with formal models of development. Lindenmayer introduced parallel rewriting systems to describe cellular growth [1], and Prusinkiewicz and Lindenmayer developed them into a general account of plant form [2]. Honda showed that a few branching angles and length ratios reproduce a wide range of tree silhouettes [14], and de Reffye et al. grounded stochastic growth in botanical

architecture [15]. Later systems traded mechanism for control, through parametric trees [3] and component-based modelling [16], or let form emerge from competition for space, in branching [17, 18] and in leaf venation [19]. Phyllotaxis, the arrangement of organs around an axis, has its own mathematical literature [20, 21, 22]. Inverse procedural modelling recovers the parameters of a known model from observed geometry [23], and Infinigen [24] shows how far a hand-written procedural library of natural objects can be taken.

Models of particular species are equally well established. Developmental L-systems are calibrated against measurements of real plants, as for maize and canola [4]; image-based methods rebuild individual plants from photographs [5] or fit parametric crop models to them [6]. More recently, a vision–language model has been trained to emit procedural plant-architecture descriptions from images [12], and language models have helped domain experts write parametric plant templates [13]. This work aims at accurate three-dimensional structure for phenotyping, simulation or rendering.

Our setting borrows from all of it and differs mainly in its target. Like image-based and inverse modelling, it starts from observations of a particular species; unlike them, there is no fixed model whose parameters are fitted, and the program itself is what the model writes and revises. Like a hand-written library, each program is imperative code for one species, here written by a language model from reference images. The target is a two-dimensional drawing that must be recognised at a glance, so we borrow the literature's botanical vocabulary, phyllotaxis, the raceme, the umbel and dichotomous venation, but none of its machinery of grammars or growth simulation.

2.2 Pen-and-ink and illustrative rendering

Line drawing from geometry is a core topic of non-photorealistic rendering [25, 26]. Winkenbach and Salesin rendered models with prioritised stroke textures [27] and parametric surfaces with strokes that follow isoparametric curves [28]. Salisbury et al. made pen-and-ink illustration interactive [29, 30], Hertzmann and Zorin laid hatching along principal curvature directions [31], and Deussen and Strothotte drew trees in pen and ink by abstracting foliage [32]. These methods shade a surface that already exists. Our programs have none: hatch direction and spacing come from the frames of the two-dimensional curves that define each organ.

2.3 Language models that write and revise graphics programs

Language models trained on code [33] write programs whose output is visual, from SVG and TikZ [7, 8, 34, 35] to procedural three-dimensional scenes [9]. Closest to our concern are systems that revise such programs after looking at their output: SceneCraft refines Blender scripts against rendered images [10], and BlenderAlchemy edits materials and scenes with a vision–language critic [11]. More generally, a model can improve its own output by critiquing it in words and trying again [36], and can carry those critiques forward as a written memory [37]. Ellis et al. inferred graphics programs from hand drawings [38], the inverse problem at the scale of simple figures.

This work mostly measures the output of a single generation or a short automatic loop against a benchmark. We instead study a recorded practice: long-lived programs refined over several passes and then kept, a critic that compares each render with images of one real species rather than with a prompt, a written memory shared across species, and, because the records survive, the defects themselves as data.

2.4 Botanical illustration practice

The visual target is the engraved and hand-coloured botanical plate of the eighteenth and nineteenth centuries [39, 40], chiefly Köhler's *Medizinal-Pflanzen* [41] and *Curtis's Botanical Magazine* [42]. From them we take the habit drawing, the enlarged detail set beside it, and colour laid as a wash over the line. The principle behind all three is that of scientific illustration generally [43]: an illustration exists to make its subject identifiable. How to evaluate illustrative rendering against that aim is itself open [44, 45], and this study does not settle it (Section 6).

3 The Study

3.1 Corpus

The corpus is every plant drawn for poisongarden.org,[1] the companion site to a book in progress on naturally occurring poisons and venoms: 33 species written over two weeks in September 2026. Each is a TypeScript program of 292 to 897 lines, median 478, 16,276 lines in all, sharing two small modules of curve and helper code, and, for the later plants, one of occlusion code, each of about 120 to 140 lines. Every line was written by language models of Anthropic's Claude family. The author, who owns the site, conceived the project, designed the render-and-revise loop, supplied and approved the references, read the code and the renders, and accepted or rejected the drawings. Toxicity plays no part in the method.

The drawings were made in three consecutive phases that differ in what the model had in hand, and the difference is the natural experiment at the centre of Section 5.1. We say a plant was drawn *from knowledge* when the model had text and no image, whether the text was its own botanical knowledge or a written description, and *from references* when it had at least one image of the species.

Phase K *Knowledge.* Nine plants, each written from a specification that a model composed from botanical knowledge alone, with no image in view.

Phase P *Plates.* Fifteen plants, thirteen of them written with one or two public-domain plates in hand; one was written from knowledge alone, and for one the record does not say.

Phase R *References.* The last nine plants, during which the protocol settled: references gathered before any code, plates and, where any exist, field photographs; a written read of

[1] https://poisongarden.org

them (Section 3.4); and a shared document of lessons, the *field notes*, read before each new plant and added to after each pass. The first of them, the Jerusalem cherry, was deliberately begun from text alone as a test and then rebuilt from references.

Between Phases P and R came the *plate review*, a single audit in which a separate model was shown renders of all 24 plants of Phases K and P, each beside its approved plate, and scored each from 1, would not be recognised, to 5, reads clearly as the species; five plants were then redrawn with their plates in hand (Section 5.3). In all, eleven plants began from knowledge and twenty-one from references; for one the record does not say.

3.2 Drawings as programs

A drawing of species s is produced by a *generator*, a program

$$G_s : \mathcal{U} \times [0.3, 1] \to \mathcal{S}, \qquad (r, d) \mapsto S, \tag{1}$$

where $r = (u_1, u_2, \dots)$ is a stream of uniform variates from a seeded pseudo-random generator [46], $\mathcal{U}$ the set of such streams, d is a level of detail, and S is a *scene*: ink strokes, each a path of line and Bézier segments with a width, an opacity and an optional tint, together with translucent colour washes and short text notes. Coordinates follow the SVG convention, y downwards, with the plant's foot at the origin. The generator reads nothing but its arguments, so a seed and a detail level always give the same drawing, byte for byte, and different seeds give different specimens of one species, as a herbarium holds many sheets of one plant. Detail enters through $\hat{d} = (d - 0.3)/0.7$, and each count it controls is interpolated between a sparse and a full value,

$$n(\hat{d}) = \operatorname{round}\big((1 - \hat{d})\, n_{\min} + \hat{d}\, n_{\max}\big). \tag{2}$$

Two budgets apply to every drawing: its width must not exceed 210 px, and at $d = 0.7$ it should carry roughly 300 to 600 strokes.

3.3 The shared kernel

The generators share very little code: a few curve functions, a tone convention and an occlusion queue. Everything else is written per species, and the case studies of Section 4 build on the constructions below.

Frames on a curve. Stems, stalks, midribs and petioles are quadratic or cubic Bézier curves,

$$\mathbf{B}_2(t) = (1-t)^2\mathbf{P}_0 + 2(1-t)t\,\mathbf{P}_1 + t^2\mathbf{P}_2, \tag{3}$$

$$\mathbf{B}_3(t) = \textstyle\sum_{i=0}^{3} \binom{3}{i}(1-t)^{3-i}t^i\,\mathbf{P}_i, \qquad t \in [0,1]. \tag{4}$$

Differentiating gives every point a unit tangent and normal,

$$\mathbf{T}(t) = \frac{\mathbf{B}'(t)}{\|\mathbf{B}'(t)\|}, \qquad \mathbf{N}(t) = \begin{pmatrix} -T_y(t) \\ T_x(t) \end{pmatrix}, \tag{5}$$

falling back to the chord direction where $\mathbf{B}'$ vanishes. This frame orients everything drawn along a curve: the two edges of a stem, the hatch across it, and the leaves and flowers attached to it.

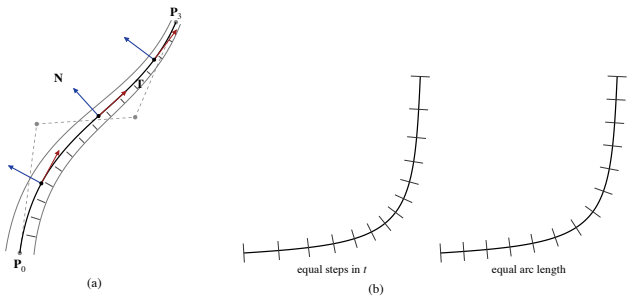


Figure 2: The curve kernel. (a) A cubic Bézier stem with its control polygon (dashed), frames $(\mathbf{T}, \mathbf{N})$ from (5) at three parameters, the two edges of (6) under a tapering width, and hatch strokes (10) on one side, placed by arc length. (b) Fourteen marks placed at equal steps of t (left), which crowd where the curve turns, and at equal arc length by (9) (right).

Ribbons and blades. A stem of width profile $w(t)$ is drawn as its two edges,

$$\mathbf{E}_{\pm}(t) = \mathbf{B}(t) \pm \tfrac{1}{2}\, w(t)\, \mathbf{N}(t), \tag{6}$$

sampled as polylines, since an offset of a Bézier curve is not itself polynomial. The edges are free of loops wherever the half-width is less than the radius of curvature,

$$\tfrac{1}{2}\, w(t) < \rho(t) = \frac{\|\mathbf{B}'(t)\|^3}{\left|\mathbf{B}'(t) \times \mathbf{B}''(t)\right|}, \tag{7}$$

with $\times$ the scalar cross product of plane vectors. A leaf blade is the same construction with the midrib as the curve and a half-width profile $h(t)$ that is zero at both ends. Writing it in the midrib's own frame,

$$\mathbf{x}(t, q) = \mathbf{B}(t) + q\, h(t)\, \mathbf{N}(t), \qquad q \in [-1, 1], \tag{8}$$

makes the outline the curves $q = \pm 1$, puts a vein at any curve $q(t)$ in between, and lets the blade bend, droop or twist by moving only the midrib's control points. This local frame is the most reused idea in the corpus.

Spacing by arc length. Marks placed at equal steps of t crowd where a curve turns tightly, so marks are spaced by arc length, $s(t) = \int_0^t \|\mathbf{B}'(\tau)\|\, d\tau$. The integral is tabulated by $K = 100$ chords, $\hat{s}_j = \sum_{i=1}^{j} \|\mathbf{B}(i/K) - \mathbf{B}((i-1)/K)\|$, with error $O(K^{-2})$, and a target length σ with $\hat{s}_j \le \sigma < \hat{s}_{j+1}$ is inverted linearly,

$$t(\sigma) = \frac{1}{K}\left(j + \frac{\sigma - \hat{s}_j}{\hat{s}_{j+1} - \hat{s}_j}\right). \tag{9}$$

Fig. 2 shows the frames, the ribbon and the difference arc length makes on a curve whose speed varies strongly.

Hatching. Engravers model form with strokes that run across a surface and follow its turn. A hatch stroke at parameter t

on side $\varsigma \in \{-1, +1\}$ runs along the normal between two fractions of the half-width,

$$\mathbf{h}(t) = \Big[\, \mathbf{x}(t, \varsigma f_{\text{in}}),\ \ \mathbf{x}(t, \varsigma f_{\text{out}}) \,\Big], \tag{10}$$

with $f_{\text{in}} = 0.25$ and $f_{\text{out}} = 0.95$, at parameters from (9). Because the direction comes from the local normal, the hatching turns with the organ, which is what curvature-aligned hatching achieves on surfaces [28, 31]. One light, from the upper left, decides which side is hatched.

Tone on a dark ground. The drawings are light ink on a near-black ground, which reverses the logic of engraving. A stroke of opacity α moves a pixel's luminance a fraction α of the way towards the ink, so hatching of stroke width w and spacing Δ gives a mean luminance of about

$$\bar{Y} \approx Y_{\text{ground}} + \frac{w}{\Delta}\,\alpha\,\big(Y_{\text{ink}} - Y_{\text{ground}}\big). \tag{11}$$

On paper hatching darkens; here it brightens. Tone is therefore raised by hatching and lowered by leaving ground bare, and a feature that should read as dark is made by *reserving* it. Colour appears only where the plant has it, as a translucent wash under or over the line, never as an opaque fill.

Depth-sorted occlusion. An engraved surface is an outline, not a filled shape, so the painter's algorithm cannot hide a far leaf behind a near one: the far leaf's veins would show through. Dense plants therefore hide ink by clipping it. Every stroke enters a queue with a depth z, growing towards the viewer, and an owner o; every surface registers its outline as a polygon Q_k with depth z_k, owner o_k and bounding box β_k. A point $\mathbf{q}$ of a stroke is hidden when

$$\text{hid}(\mathbf{q}) \iff \exists\, k :\ z_k > z \ \wedge\ o_k \neq o \ \wedge\ \mathbf{q} \in \beta_k \ \wedge\ \mathbf{q} \in Q_k, \tag{12}$$

with membership in Q_k decided by the even–odd rule. Strokes are sampled every $h = 1.2$ px and each maximal run of visible samples is kept, so a hidden span is located to within h. Siblings are given depths $z_0 + k\varepsilon$ so that equal depths never tie, and a surface that is mostly covered is demoted to outline and midrib at lower opacity. The queue was first written for the oleander, rewritten for the Jerusalem cherry (Section 4.2), copied into bloodroot, and made a shared module for the plants after it.

3.4 The refinement loop

Four roles take part (Fig. 3). A *builder*, a model working under the lead's instructions, writes the program, starting from a *donor*, the program of an earlier plant of similar build. A *lead*, a separate model session, gathers references, writes the brief and looks at every contact sheet, a grid of renders across seeds. A *reviewer*, a model with no stake in the code, compares drawings with plates or reviews a change before it merges. The *owner*, the author, designed the loop, reads the code and the renders, directs revisions, and has the final word.

The lead first turns the references $\mathcal{R}_s$ into a *read*,

$$\rho_s = \big(\mathbf{t},\ \mathcal{A},\ o^{\star}\big), \tag{13}$$

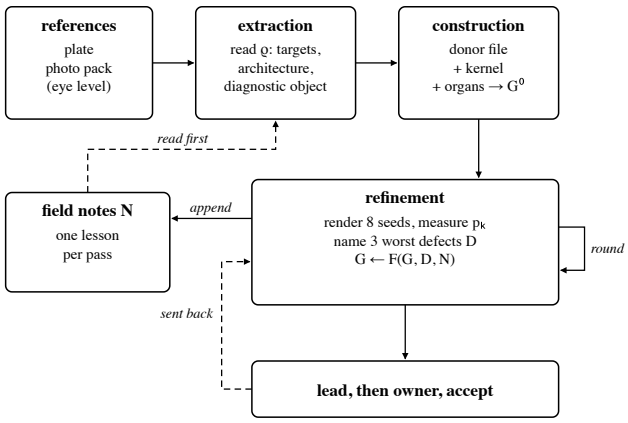


Figure 3: The loop. References are read into a written read; a builder writes a first program from a donor file, a shared kernel and organ constructions; each round renders eight seeds, names the three worst defects and revises the program; the lead and then the owner accept the drawing or send it back; one lesson per pass goes into shared field notes, which every later brief reads first.

where $\mathbf{t} = \{(m_k, [\ell_k, u_k])\}$ are numeric targets, each a measurement m_k computable on a scene, such as the ratio of width to height, with an interval it should fall in; $\mathcal{A}$ is the architecture, what rises from what and how many; and $o^{\star}$ is the one diagnostic object that must read above all others. Round i of a pass renders the current program on a sample of seeds Σ, eight in practice, computes the fraction of seeds meeting each target,

$$p_k\big(G^{(i)}\big) = \frac{1}{|\Sigma|}\sum_{r \in \Sigma} \mathbf{1}\big[\ell_k \le m_k\big(G^{(i)}(r, d)\big) \le u_k\big], \tag{14}$$

sets the renders beside the references with key parts cropped at three times their size, and writes down the three worst defects $D^{(i)} = \{\delta_1, \delta_2, \delta_3\}$ before changing any code. The revision

$$G^{(i+1)} = \mathcal{F}\big(G^{(i)},\, D^{(i)},\, \mathcal{N}\big) \tag{15}$$

is the model's rewrite of its program in the light of the defects and of the field notes $\mathcal{N}$. Rounds are capped, at six for a first pass and two for a polish pass, and a pass ends with a list of what is still weak rather than with more rounds. This is verbal self-refinement [36] with a persistent written memory [37], except that the critic compares the output with images of one real species, and the memory is shared across species.

No differentiable proxy for recognisability exists, so the loop cannot be replaced by optimisation of a fixed construction, as in inverse procedural modelling [23] or optimisation against an image prior [47]. The construction is not fixed: choosing it, as the case studies show, is most of the work.

3.5 Data and coding

For each plant we assembled every version of its program from the repository (61 versions), its written read where one

Table 1: The five case studies, in the order drawn, with the phase of the corpus each belongs to. Tokens are the model usage recorded in the project's cost log, approximate for the last two; none was kept for the foxglove.

Plant	Phase	In hand	Passes	Tokens
Foxglove	K	knowledge; plates for the redraw	1 + redraw	–
Jerusalem cherry	R	text; a photo, then a plate	4	434,128
Calabar bean	R	two plates	2	365,796
Bloodroot	R	a plate, 4 photos; 6 more	3	407,000
Sago palm	R	30 photos, a read, 2 plates	1	150,000

exists, the briefs and reviews, the shared field notes, and the transcripts of the sessions that wrote it, including the builders' own. The completeness of the record varies. For 21 plants the defects of individual rounds survive; for 10 only a summary per pass; for 2 only the commit messages. The plate review gives a common yardstick for the plants of Phases K and P.

From these records we extracted every named defect that led to a change or was explicitly left open: 221 in all. Each was coded on three dimensions. Its *kind* is one of eight: habit, organ form, presentation, density and occlusion, legibility and scale, tone and colour, budget, or not a drawing defect (a crash, a measuring bug). It was *found* either by *measurement*, a number outside a numeric target, or by *sight*, anything else. Its *fix* was *structural*, replacing a construction; *parametric*, changing only constants; or *open*. We also recorded who first named it. The extraction and the first coding were made per plant from the records. Kind and found were then coded again by two independent coders, both language-model instances working blind from the defect text alone, and fix type was coded a second time from the source commits and transcripts for a random sample of 40 fixed defects. Agreement is reported with Cohen's κ,

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{16}$$

where p_o is the observed proportion of agreement and p_e the agreement expected from the coders' marginal frequencies. Disagreements were adjudicated by a third reading; for fix type, the code checked against the source was kept. The codebook, the coded defects and the per-plant records are part of the supplementary material.

4 Five Case Studies

The five plants are taken in the order they were drawn, which is also the order in which the protocol took shape. For each we follow the model's own record: what it had in hand, what it wrote down, what it built first, what it named as wrong, and the construction that replaced the first. Table 1 summarises them and Fig. 4 shows the versions. Together the five answer RQ4: the constructions that carried recognisability are named in each final construction, and Section 6.1 draws them together.

4.1 Foxglove: the textbook rule

In hand. Nothing but knowledge. At the start of Phase K the lead model wrote a specification for the first three plants from botanical facts alone, and a builder implemented it. For foxglove it asked for bells "arranged spirally: place flower k at … $x = r\cos(k \cdot 137.508°)$ so the spiral phyllotaxis reads in 2D".

First construction. The builder placed bell k at height h_k and lateral offset

$$x_k = x_0 + r\cos(k\psi), \qquad \psi = 2\pi\big(1 - \Phi^{-1}\big) \approx 137.5°, \tag{17}$$

with Φ the golden ratio and $r = 22$ px, and drew the bells with $\cos(k\psi) < 0$ at lower opacity, as the far side of a spiral. The golden angle is the divergence angle of spiral phyllotaxis [20, 21]. The rule is botanically defensible and pictorially wrong: it puts bells on both sides of the stem.

Defects. The plate review scored the drawing 2 of 5. Its row read: "about 19 small hollow bells wind round the stem (spiral radius 22), so they alternate sides … The plate has a dense one-sided spike of large, overlapping tubular bells." It added that the bells had no spotted throat, the top fifth were bare ovals, and the rosette pushed the width to 236 px, over the budget.

Final construction. Whatever the arrangement of the flower stalks, the flowers of a foxglove hang to one side, and that is what a plate shows and a reader recognises. The redraw after the plate review, made by a builder with two plates in hand, puts every bell on one side $s \in \{-1, +1\}$, chosen once per specimen, at stem fractions

$$t_k = t_0 + (t_1 - t_0)\big(1 - (1 - u_k)^{1.3}\big), \qquad u_k = \frac{k}{n-1}, \tag{18}$$

with $t_0 = 0.41$ and $t_1 = 0.955$. The map has slope 1.3 at the bottom and zero at the top, so the bells crowd towards the apex into the dense spike of the plate. Bell length falls as $L = L_{\max}(1 - 0.16\,u/0.66)$ for $u \le 0.66$; above that length and the opening of the mouth fall together until, beyond $u = 0.82$, the flowers are closed buds. Each bell is a swollen tube in the frame of its own axis, its mouth cut obliquely (upper rim at $t = 0.94$, lower lip at $t = 1.06$), with three to five spots inside it.

What it shows. The model's first rule described how the plant is built; the plant is known by how it presents itself. The same golden angle reappears, correctly, in the rays of the hemlock umbel and in the crown of the sago palm below, where it spreads organs evenly round an axis without a visible grid.

4.2 Jerusalem cherry: from text to a bush

In hand. Text only. The Jerusalem cherry opened Phase R as a deliberate test: the lead commissioned it to be drawn from the published article, with no image, and its cost counted.

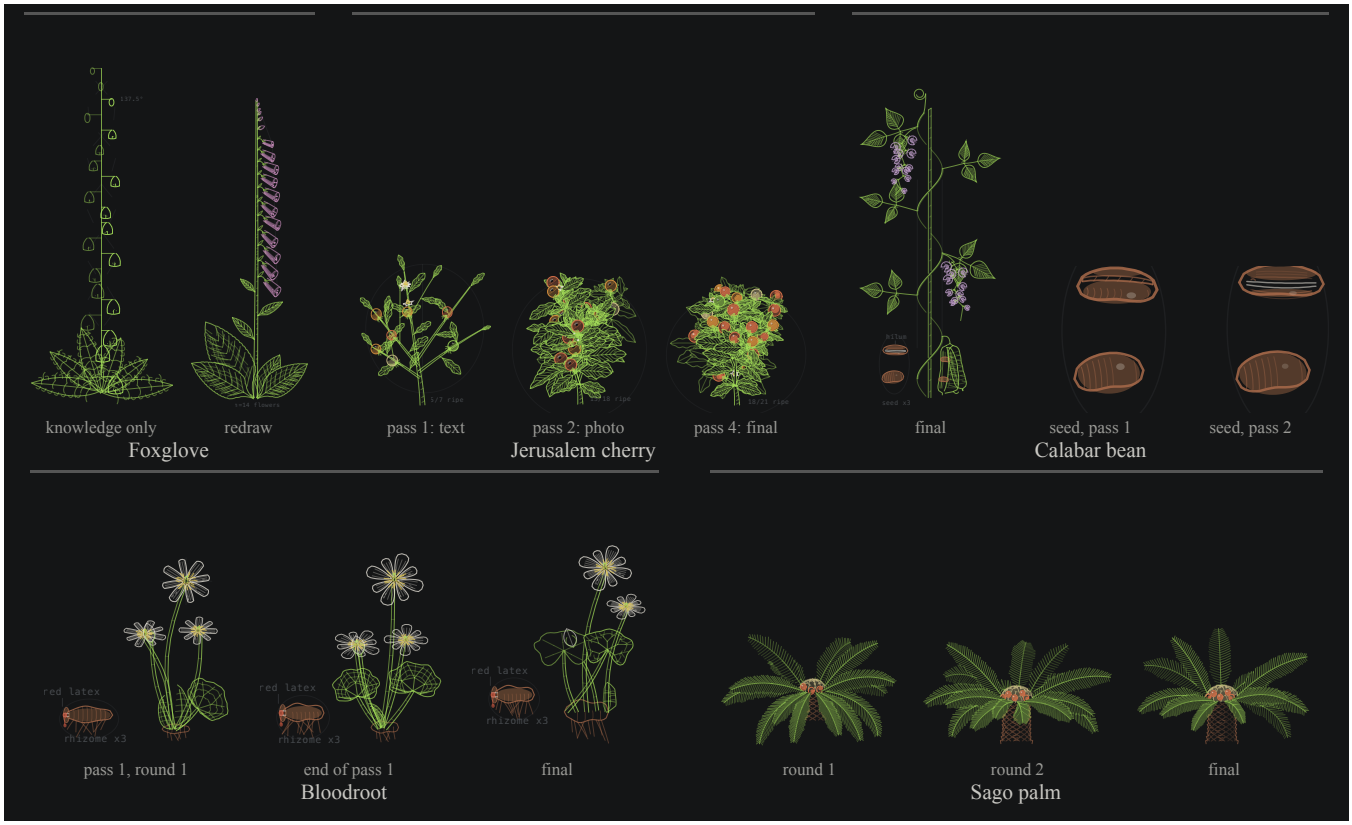


Figure 4: The case-study plants from first to final program. Foxglove: drawn from knowledge alone, and redrawn against its plate. Jerusalem cherry: the four committed passes, from a text-only fan to the final dome. Calabar bean: the plant, with the enlarged seed from pass 1 and pass 2. Bloodroot: pass 1 round 1, the end of pass 1, and the final drawing. Sago palm: rounds 1 and 2 and the final drawing. Seed 11 and detail 0.9 throughout, one scale within each plant. Intermediate rounds of bloodroot and the sago palm were rebuilt by replaying the builder's recorded edits; both replays reproduce the committed code exactly.

First construction. The builder wrote a short trunk carrying five primaries fanned from points spaced up it. Primary k, $u = k/4$, leaves at angle

$$\varphi_k = \sigma_k\big(112^\circ - 88^\circ u + \epsilon_k\big), \qquad \epsilon_k \in [-8^\circ, 8^\circ], \tag{19}$$

from vertical, alternating side σ_k, and runs as one quadratic curve to a point on a canopy ellipse; fruit and flowers take fifteen fixed sites along the primaries. There is no depth: every outline is drawn in a fixed order. After six rounds the builder's own verdict was "a stiff radiating fan/rosette, not a rounded woody bush". The owner's was blunter: "It is nowhere near dense enough. The drawn one currently looks like a sick and dying version of 'maybe' a jerusalem cherry plant."

The read. The owner supplied a photograph of a healthy plant, and the lead turned it into numbers before any code changed: "a dense, full, rounded dome, a little WIDER than tall, leafy right down to the base ... If I can trace a branch from trunk to tip, it is wrong"; 70 to 120 leaves; 14 to 24 berries of diameter 0.4 to 0.5 of a leaf's length; a box of 190–205 px wide. For the third pass a plate arrived, Van Houtte's of 1857, under the old name *Solanum capsicastrum*.

Second construction: growth under an envelope. Pass 2 replaced the fan with recursive growth inside a squat crown ellipse of centre $\mathbf{C}$ and semi-axes $r_x \in [98, 102]$, $r_y \in [84, 92]$. A shoot is a walk of nodes $\mathbf{p}_i$ with heading h_i. With $\mathbf{d} = \big((p_x - C_x)/r_x,\ (p_y - C_y)/r_y\big)$ and $\rho = \|\mathbf{d}\|$, the shoot is steered by

$$\mathbf{w} = \operatorname{unit}\big(\lambda(\rho)\, d_x,\ \lambda(\rho)\, d_y - 0.38\big), \qquad \lambda(\rho) = \begin{cases} 0.34, & \rho \le 0.88, \\ -5(\rho - 0.88), & \rho > 0.88, \end{cases} \tag{20}$$

outward and upward inside the envelope and sharply back near its surface, and it kinks at every node:

$$h_{i+1} = h_i + \tfrac{1}{2}\,\Delta\big(h_i, \angle\mathbf{w}\big) + (-1)^i\, \varepsilon\, \eta_i, \qquad \eta_i \in [15^\circ, 27^\circ], \tag{21}$$

where Δ is the signed angular difference and $\varepsilon = \pm 1$ a parity fixed per shoot. A shoot stops if it would leave the envelope, $\rho > 1.02$, and forks up to order three, two daughters from a primary limb and one or two from each later shoot, each leaving at 34° to 62°. The alternating kink is what makes the lead's test pass: no branch can be traced from trunk to tip. Every leaf and berry now entered the depth queue of (12).

Defect: the berry as a hole. The bush was right; the fruit was not. In the builder's words: "each berry punches a black hole in the leaf mass — on a dark ground with light ink, my shading gradient is inverted. Ink means light here." By (11), every stroke raises luminance, and an unfilled outline that hides the foliage behind it shows bare ground. The fix inverted the engraver's habit, with dense contour arcs on the lit side thinning towards the shadow and a reserved window, and added a translucent wash beneath each berry's ink. A partly hidden berry cannot take a full disc, and an SVG ellipse cannot be clipped, so the wash is the largest circle that fits in the berry's visible part. With visible and hidden samples V and H inside the ring of centre $\mathbf{c}$ and radius r,

$$\mathbf{p}^{*} = \arg\max_{\mathbf{p}\in V} \min\Big(0.94\,r - \|\mathbf{p}-\mathbf{c}\|,\ \min_{\mathbf{q}\in H}\|\mathbf{p}-\mathbf{q}\|\Big), \tag{22}$$

and the inner minimum at $\mathbf{p}^{*}$ is the wash radius; a berry with less than $0.26\,r$ of clear radius gets none. The final pass, at the owner's request, re-centred the washes, which had been nudged towards the shadow and made the berries read as eyes. Measured on the render, the berries' mean luminance rose from 91, below the foliage's 96, to 106.8, above it.

What it shows. Four passes, about twenty rounds and 434,128 tokens, against 365,796 for the Calabar bean below, drawn next, from plates alone. The knowledge-only draft was structurally wrong, and no round of tuning could have made a fan into a bush.

4.3 Calabar bean: plates, and a reserved groove

In hand. Two plates, Köhler (1890) and Berg and Schmidt (1896), and nothing else: no photograph of the living plant could be found on iNaturalist, only herbarium sheets. The plates disagreed about colour. Köhler prints the flowers coral red; the second plate and every written description say pale pink to purple. The disagreement was caught and settled from the text before any code, and the drawing follows the text.

The read. The builder's header comment is its read: a slender liane climbing by twining, never by tendrils; few, very large trifoliate leaves; hanging racemes of flowers "each . . . a snail-shell coil"; flat pods, one split to show the beans; and the bean itself enlarged beside the foot, "with the black groove running end to end DOWN THE FACE between two pale raised lips".

Construction. The stem twines as a helix about a dead cane, $x(y) = x_c(y) + a\sin\theta(y)$ with θ linear in height, its depth given by $\cos\theta(y)$, so passes in front cut the cane's outline and passes behind are hidden. The keel of each flower is a logarithmic spiral, $r(\theta) = r_0 e^{-k\theta}$ with $k = \ln 5/3\pi$. Pass 1 took six rounds and landed near its final form; its round-by-round defects were a measuring bug, pods like "lumpy sausages", racemes like a "fishbone ladder", and a split pod that read as an x-ray.

Defect: a dark groove on a dark ground. The owner accepted the plant and asked for one more pass on the seed: the groove "reads muddled at 3-4x and vanishes at garden scale", that is, at page size. A dark feature cannot be inked on this ground, since ink only adds light; it has to be *reserved*. The seed of half-length a and half-width b takes a brown wash in two bands either side of a channel of half-height $g = 0.13\,b$ that sags along the face,

$$\begin{aligned} \mathrm{bow}(t) &= 0.12\,b\,\sin(\pi t),\\ q^{\pm}_{\mathrm{lip}}(t) &= \pm g\,\sin^{0.32}(\pi t) + \mathrm{bow}(t), \end{aligned} \tag{23}$$

and two pale lips, the brightest lines in the seed, run along the channel and pinch to a point at each end. Pass 1 had placed the groove along the convex edge rather than down the face, and in pass 2 the builder found a second, geometric cause the brief had not named. It wrote a probe that measured each wash against the seed's outline and found bands crossing it by up to 1.4 px, because the seed was drawn slightly tilted and an SVG `<ellipse>` has no rotation: a thin axis-aligned band slides out past a tilted outline at its ends. The fix levelled the seed and fitted each band inside the outline. Inverting the outline profile $b\,\sin^{0.42}(\pi t)$ gives the half-length available at an offset q across the seed,

$$\ell(q) = a\Big(1 - \tfrac{2}{\pi}\arcsin\big(\min(1, |q|/b)^{1/0.42}\big)\Big), \tag{24}$$

with b replaced by $0.82\,b$ on the flatter face side, and each band's semi-axis along the seed is the least of ℓ over the band's height, sampled at eight angles.

What it shows. Plates alone, read carefully and cross-checked against text, gave a near-final drawing in one pass. The one defect the owner raised was a matter of tone on a dark ground, and a second cause, invisible at page size, was found by measurement, not by eye.

4.4 Bloodroot: photographs from the wrong angle

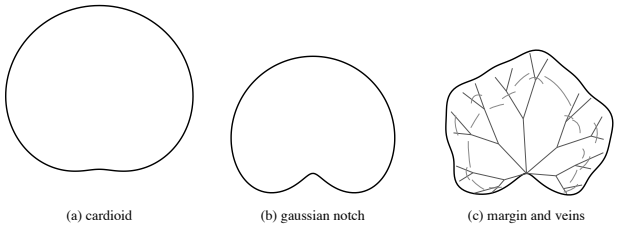
In hand. A plate and four photographs: a flower from above, a leaf from above, a clump, and a shoot. Not one was taken at eye level.

The geometry of views. Under orthographic projection a planar organ keeps its extent along any axis in the picture plane and loses it along any axis that tilts away. A leaf blade held horizontal and seen from an elevation ε keeps its width and shows its depth multiplied by

$$f(\varepsilon) = \sin\varepsilon, \tag{25}$$

so a photograph from above, $\varepsilon = 90°$, shows the true outline, and one at eye level shows it nearly edge-on. Proportions read from the one cannot be carried to the other. The read asked for flower stalks one and a half to two times the height of the leaves; the first pass drew them, and the plant read, in the lead's words, as "a bouquet of daisies".

First constructions. The blade was a cardioid, $r(\varphi) = 1 - k\cos\varphi$, the standard closed curve with one notch. Its single parameter sets the notch's depth, but the cosine spreads the notch over the whole circle, so a deep notch cannot also be



(a) cardioid (b) gaussian notch (c) margin and veins

Figure 5: The bloodroot blade, stalk at the bottom. (a) A cardioid, $r = 1 - 0.62\cos\varphi$: a deep notch cannot also be narrow, and the blade reads as a rounded square. (b) A gaussian notch (26), $\delta = 0.57$, $\omega = 0.58$: two rounded basal lobes. (c) The full margin (27) with dichotomous venation, 54 strokes. Computed with the generator's own formulas.

narrow: at $k = 0.62$ the basal lobes sat 0.8 px below the notch and the blade read as a rounded square. The veins were rays from the notch crossed by rungs at shared radii, the regular net of a textbook diagram.

Refinement. In round 3 the builder bit a gaussian notch out of a disc,

$$n(\varphi) = \delta\,\exp\!\big(-(\varphi/\omega)^2\big), \qquad \varphi \in (-\pi, \pi], \tag{26}$$

which gives depth δ and width ω separate controls, and carried lobes and teeth on the margin, damped inside the notch so that the sinus stays clean,

$$\begin{aligned} r(\varphi) = \big(1 - n(\varphi)\big)\big(1 + a_L\,\kappa(\varphi)\cos(m\varphi + \theta_L)\big) \\ \times \big(1 + a_T\,\kappa(\varphi)\cos((2m+1)\varphi + \theta_T)\big), \end{aligned} \tag{27}$$

with $\kappa(\varphi) = 1 - n(\varphi)/\delta$ (Fig. 5). After six rounds the lead sent the pass back with five defects, the first "the scapes are far too long, too straight, and all leave one point". The owner ruled at the same time that reference photographs, which are never published, need not be openly licensed, and pass 2 found six eye-level photographs. One side-on shoot fixed the stalk length, the height of the leaf and the size of flower against leaf at once: the flower clears its own leaf by about one flower width. The height moved from the flower stalks to the leaf stalks; the shoots rose from slots spread along the rhizome and shuffled, so that the kind of shoot and its place are independent; the blades became horizontal plates, their depth scaled by $0.84\,\eta$ with $\eta \in [0.55, 0.75]$, which by (25) is a view from $\varepsilon = \arcsin(0.84\eta) \in [27.5°, 39.1°]$, as a person standing over a woodland plant sees it. The veins became dichotomous, as real venation is [19]: five primaries forking twice, the twenty tips joined by nineteen cross-links, 54 strokes at every detail, about half the strokes of the web they replaced. Pass 3 added the occlusion queue to clear a tangle at the foot, which then exposed a furled young leaf drawn far too large.

What it shows. References are not enough; they must be taken from the view the drawing needs. Every round of bloodroot survives, and it is the one plant in which the builder's and the lead's judgements can be compared round by round (Section 5.3).

4.5 Sago palm: the settled protocol

In hand. A pack of thirty photographs with a written read, and two plates from *Curtis's Botanical Magazine*, one of the habit and one of the female cone. The read gave the silhouette as numbers, W/H about 1.3 to 1.7 and a trunk of 15 to 35 per cent of the height, and named what a reader must see: stiff arching fronds, a trunk armoured with old leaf bases, and in a female plant the domed cone with its orange seeds, the most toxic part. No earlier plant was close enough to serve as a donor.

First construction. The first draft already had its final architecture. Each frond is a three-dimensional rachis at golden-angle azimuth φ_i, starting at an elevation e_0 that falls with the frond's age a and bending down towards its tip, stepped in arc length Δs as

$$\begin{aligned} \mathbf{p}_{k+1} = \mathbf{p}_k + \Delta s\,\big(\cos e_k \sin\varphi_i,\ \sin e_k,\ \cos e_k\cos\varphi_i\big), \\ e_k = e_0 - \beta\Big(\tfrac{k+1/2}{n}\Big)^{1.6}, \end{aligned} \tag{28}$$

and projected by an orthographic camera pitched down by γ,

$$x_s = x_0 + x, \qquad y_s = y_0 - (y\cos\gamma - z\sin\gamma). \tag{29}$$

The leaflets are not placed in three dimensions. They are laid in screen space, at 56° forward of the projected rachis's tangent, and the program's header gives the reason: "so every frond presents its comb (the Curtis view) rather than going edge-on". That is the lesson the field notes had recorded from earlier plants, that a leaf seen edge-on never reads, arriving as a design choice. The leaflets also serve as the frond's hatching.

Refinement. Four rounds, each with the defects named in the builder's preserved reasoning. Round 1: "fronds cover the trunk and hug the ground making the plant look flat, the seeds sit on the cone's rim like beads instead of on a tan dome". The camera came down from $\gamma = 14°$ to 10°, the trunk lengthened, and every rachis point was clamped to the ground. Solving $y_s = y_0 - 5$ in (29) for world height gives the clamp

$$y \geq \frac{5 + z\sin\gamma}{\cos\gamma}, \tag{30}$$

so an old frond runs along the ground line instead of through it. Fronds facing the viewer, $\cos\varphi_i > 0.3$, start lower and arch less, and their leaflets shorten with the frond's projected length, which removed "a dense green bib under the cone". The trunk's armour is two families of helices about its axis, $x = x_a(y) + r(y)\sin\theta$, $y = y_0 \pm k\theta$ with $k = r_0/1.7$, drawn at opacity 0.32 on the lit side and 0.22 on the shaded; at 0.38 it "read as fishnet". The code review found nothing to change, and six weaknesses were recorded as open.

What it shows. By the last plant the loop had become cheaper and quieter: one pass, four rounds, about 150,000 tokens, no

rejection. The lessons of earlier plants, about presentation, occlusion and washes, arrived as design choices in the first draft rather than as defects in later rounds.

5 Results Across the Corpus

5.1 Grounding (RQ1)

The plate review gives a common score for the 24 plants of Phases K and P. Of these, 10 were first written from knowledge alone, the nine of Phase K and one of Phase P, and 13 from plates; for one the record does not say, and it is left out. The knowledge-only drawings scored a median of 3 against 4 (Fig. 6a). For two samples of sizes n_1 and n_2, here the knowledge-only scores a_i and the plate scores b_j, the Mann–Whitney statistic counts, over all pairs, how often a value of the first exceeds one of the second,

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Big(\mathbf{1}[a_i > b_j] + \tfrac{1}{2}\,\mathbf{1}[a_i = b_j] \Big), \qquad r = 1 - \frac{2U}{n_1 n_2}, \tag{31}$$

and r is the rank-biserial effect size. Here $U = 36$, $r = 0.45$, a medium effect, with $p = 0.03$ for the one-sided hypothesis that references raise the score and $p = 0.06$ two-sided. The width budget tells the same story more starkly. Of the first committed versions of all 32 plants whose basis is known, three exceeded 210 px at some level of detail, foxglove at 240 at its widest seed (the review had measured 236), hemlock at 230 and thorn apple at 219, and all three were drawn from knowledge alone: 3 of 11 against 0 of 21 (Fisher's exact test, $p = 0.03$; Fig. 6b). The knowledge-only programs also let detail multiply their stroke counts more, a median ratio of 1.54 between the fullest and sparsest drawing against 1.29 (Mann–Whitney $p = 0.03$). These are three views of one difference, not three independent confirmations. In the extreme case, the first hemlock's leaves scaled two counts with detail at once, the pinnae and the leaflets on each, so that a leaf carried

$$N(\hat{d}) = 1 + n_P(\hat{d})\,\big(1 + 2\,n_L(\hat{d})\big) \tag{32}$$

strokes, quadratic in $\hat{d}$ by (2); its path count grew 6.7-fold across the range of detail.

These comparisons are observational. Most knowledge-only plants were written first, in Phase K, and the reference-based ones later, in Phases P and R; the models, the experience accumulated in the sessions, the reviewer's brief and the plants themselves differ. The one within-plant comparison, the Jerusalem cherry, points the same way: its text-only pass was rejected outright and its rebuilding from references took three more passes.

5.2 Failures (RQ2)

The records name 221 defects, of which 11 are not drawing defects (crashes, bugs in measuring tools, notes about process), leaving 210. Coding agreement was substantial to almost perfect: Cohen's $\kappa = 0.88$ for kind between the two blind coders, 0.66 for found, and 0.66 for fix type on the sample of 40 (85 per cent raw agreement). Fig. 7 gives the distribution.

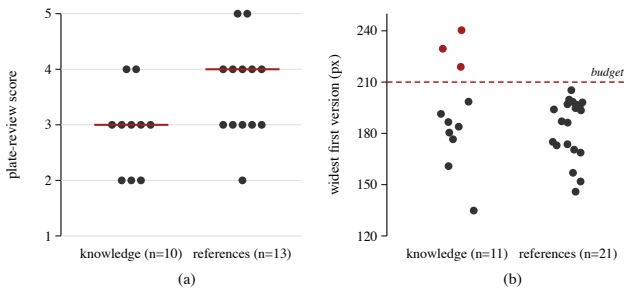


Figure 6: Grounding. (a) Score in the plate review, 1 (would not be recognised) to 5 (reads clearly), for plants first written from knowledge alone and from plates; bars are medians. (b) Widest drawing of the first committed version at any level of detail; the three over the 210 px budget were all written from knowledge alone.

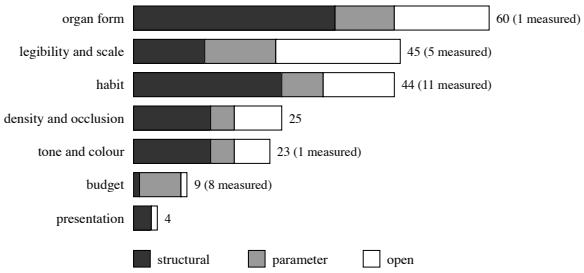


Figure 7: The 210 drawing defects named across the 33 plants, by kind and by fate: fixed by replacing a construction, fixed by changing its parameters, or left open. The count found by measurement, rather than by sight, is given in brackets.

Three findings stand out. First, the failures are about form, not geometry. Organ form, habit and legibility account for 149 of the 210, and defects in the curve geometry itself, frames, offsets and arc-length spacing, are rare in the record. Second, most defects were found by sight: 184 of 210, 88 per cent. Only nine concern the budgets, and the measured defects cluster there and in habit, where the read supplies a width-to-height target. Third, how a defect was found predicts how it was fixed. Of the fixed defects found by sight, 93 of 123, 76 per cent, were fixed structurally; of those found by measurement, 8 of 22, 36 per cent. The odds of a structural fix are 5.4 times higher for a seen defect (Fisher's exact test, $p < 0.001$). A number outside a target is usually brought inside by tuning; a drawing that reads as the wrong thing usually needs a different construction. The case studies show why: no setting of k turns a cardioid into a kidney, and no spread of five straight primaries makes a bush.

Sixty-five defects were left open. They are the recorded residue of the round caps, the "still weak" lists with which a pass ends, and, above all, the plate review. Fifty of the 51 defects named by reviewers came from it, and 37 of those

concern plants that were not redrawn; they remain in the programs today.

5.3 Critique (RQ3)

The builder named half of the defects, 104; the lead 53; a reviewer model 51; and the owner 2. The builder's defects were mostly fixed, 77 of 104, and the lead's almost all, 52 of 53, because a lead names defects when it sends a pass back; the reviewer's, 14 of 51, since most of its plants were not redrawn. The owner's two were decisive for their plants: the rejection of the Jerusalem cherry's text-only pass, and the berries that read as eyes.

Counts do not show which defects mattered, but the bloodroot record does, round by round. Within each pass the builder named and fixed most of its defects, including the replacement of the cardioid. At each pass boundary the lead, which had not written the code, sent the pass back with defects the builder had missed or underrated: the builder closed its first pass calling the clump "slightly bouquet-ish", and the lead put the bouquet first and added the edge-on blades and the spider's web. For the early plants the same role fell to the plate review: the four plants it scored 2, and hemlock, were the ones redrawn. A model reviewing its own render is a useful critic within a round, as verbal self-refinement suggests [36]; the judgement that a drawing is done came, in this corpus, from a critic that had not made it.

6 Discussion

6.1 Textbook forms and the pull towards the typical (RQ4)

Every first construction that failed in the case studies was a form a textbook gives: the phyllotactic spiral, the fan of branches, the cardioid, the net of rays and rings. Each describes how the plant is built, and none how it presents itself to someone looking at it. The foxglove makes the point most sharply, because the golden angle that failed there was right for the hemlock umbel and the sago palm's crown: the rule was sound, and applying it to appearance was not. What the references supplied, in every case, was the species' departure from the typical: the one-sided spike, the leaf held as a horizontal plate, the irregular net, the groove down the face of the seed. We saw the same pull in a method the project tried and abandoned: a shared library of generic plant parts produced drawings that were weaker exactly where a generic part replaced a bespoke one. A diagnostic feature is by definition where a species departs from the typical, which is where any strong prior is most inclined to regularise it away. A language model's first guess is such a prior, and the loop is what removes it.

6.2 Seeing and measuring

Measured defects were the minority and the easy ones: the model computes them, checks its own fix, and usually tunes a constant. Seen defects were the majority and the ones that decided recognisability: of the 104 defects named by the lead and the reviewer, the critics who sent passes back, 96 were seen, and most needed a new construction. This suggests where automation will and will not help. Pass rates (14) and budget sweeps caught the numbers reliably; nothing in the loop could compute that a clump read as a bouquet. Better automatic critics, for instance a vision–language model scoring renders against plates, are an obvious next step, but the evidence here is that the decisive judgements came from a critic that had not made the drawing, and we would expect an automatic critic to need the same independence.

6.3 Lessons as rules

In Phase R each pass ended with one lesson written into the field notes, and many of these are small pieces of mathematics: a notch must be local in angle (26); a leaf's drawn ratio is not its true ratio (25); ink on a dark ground only adds light (11), so a dark feature must be reserved; detail should scale one count at a time (32); ink should be hidden, never shortened (12). They transferred. The sago palm's first draft laid its leaflets in screen space so that no frond goes edge-on, and used the washes and occlusion that earlier plants had had to discover; its cost was about a third of the Jerusalem cherry's. This is the written memory of Reflexion [37], kept across tasks rather than across attempts, and it is portable: the notes are prose any later model can read.

6.4 Limitations

One corpus. The 33 plants were drawn for one site, in one house style, with one owner whose eye was final, by models of one family over two weeks. The phases that give the natural comparison also differ in the kernel, the models and the protocol, so the grounding results are observational.

Uneven records and model coders. Round-level defects survive for 21 plants, pass-level summaries for 10 and commit messages for 2, so defect counts partly reflect how much was written down. The extraction and every coding were made by language models; agreement was substantial, but no human coded the corpus.

The plate review. The yardstick for RQ1 is one model's scores, and its brief listed weak spots already suspected in some drawings, among them the foxglove. It is a consistent comparison across plants, not an independent measure of recognisability, which would need expert raters this study does not have.

Reproducibility. Every version in this paper can be regenerated byte for byte from its source, and the intermediate rounds we show were rebuilt by replaying recorded edits. The process is not reproducible in the same sense: the same brief given twice will not produce the same program.

Work in progress. The drawings are still being developed. Several plants still lack organs such as fruit or seeds, and the programs described here will keep changing; the results describe the corpus as it stood when this paper was written.

6.5 Safety and disclosure

The drawings illustrate poisonous plants and are not identification guides; they should not be used to identify plants, and nothing in them, or in this paper, should be used to decide whether a plant is safe to eat or handle. Every generator was written by language models, and this manuscript was drafted with the assistance of one; both uses are identified in the acknowledgments. The author has reviewed the code, measurements and text and takes responsibility for them.

7 Conclusion

Language models can write procedural programs that draw particular plant species, and revise them from reference images until an owner accepts them and a review against plates finds them readable; whether botanists would agree is untested. Across 33 plants, drawing from references rather than knowledge gave better first programs; most failures were found by sight, not measurement, and fixed by replacing a construction rather than tuning it; and where the record shows it, the defects that sent a pass back were named by a critic that had not written the code. The model's geometry was rarely the problem. Where its first guesses failed, they were textbook forms, true to how a plant is built and false to how it looks, and the work of refinement was to replace them, one construction at a time, with the departures from the typical that the references showed.

Acknowledgments

The author conceived the project, designed the render-and-revise loop, supplied and approved the references, reviewed the code and the renders, and directed the revisions. The 33 generators, the shared drawing code and the measurement scripts were written by large language models from Anthropic (Claude Opus, Sonnet and Fable models) under the loop of Section 3.4; code review of changes to the site was carried out by a model from a second vendor. The per-plant records were extracted from the project's repository and transcripts by Claude Sonnet models, and the two blind codings and the fix-type coding were made by separate Claude Sonnet instances. This manuscript was drafted with the assistance of Anthropic's Claude Opus 5.5, checked against the records by Claude Fable 5.1, and revised by the author.

References

[1] A. Lindenmayer, "Mathematical models for cellular interactions in development I. Filaments with one-sided inputs," *Journal of Theoretical Biology*, vol. 18, no. 3, pp. 280–299, 1968.

[2] P. Prusinkiewicz and A. Lindenmayer, *The Algorithmic Beauty of Plants*. New York: Springer-Verlag, 1990.

[3] J. Weber and J. Penn, "Creation and rendering of realistic trees," in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*, 1995, pp. 119–128.

[4] M. Cieslak, N. Khan, P. Ferraro, R. Soolanayakanahally, S. J. Robinson, I. Parkin, I. McQuillan, and P. Prusinkiewicz, "L-system models for image-based phenomics: case studies of maize and canola," *in silico Plants*, vol. 4, no. 1, p. diab039, 2021.

[5] L. Quan, P. Tan, G. Zeng, L. Yuan, J. Wang, and S. B. Kang, "Image-based plant modeling," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 599–604, 2006.

[6] A. J. Zhai, X. Wang, K. Li, Z. Jiang, J. Zhou, S. Wang, Z. Jin, K. Guan, and S. Wang, "CropCraft: Complete structural characterization of crop plants from images," 2024, arXiv:2411.09693; International Conference on 3D Vision (3DV) 2026.

[7] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, "Sparks of artificial general intelligence: Early experiments with GPT-4," 2023.

[8] J. A. Rodriguez, S. Agarwal, I. H. Laradji, P. Rodriguez, D. Vázquez, C. Pal, and M. Pedersoli, "StarVector: Generating scalable vector graphics code from images and text," 2023.

[9] C. Sun, J. Han, W. Deng, X. Wang, Z. Qin, and S. Gould, "3D-GPT: Procedural 3D modeling with large language models," 2023.

[10] Z. Hu, A. Iscen, A. Jain, T. Kipf, Y. Yue, D. A. Ross, C. Schmid, and A. Fathi, "SceneCraft: An LLM agent for synthesizing 3D scene as Blender code," 2024.

[11] I. Huang, G. Yang, and L. Guibas, "BlenderAlchemy: Editing 3D graphics with vision-language models," 2024.

[12] H. Yun, I. K. Uyehara, I. Droutsas, E. Ranario, C. H. Diepenbrock, B. N. Bailey, and J. M. Earles, "A vision language model for generating procedural plant architecture representations from simulated images," 2026, arXiv:2603.22622.

[13] M. Hadadi, T. Z. Jubery, P. S. Schnable, A. Singh, B. Benes, A. Krishnamurthy, and B. Ganapathysubramanian, "FloraForge: Procedural generation of editable and analysis-ready 3D plant geometric models using LLM-assisted template design," *Smart Agricultural Technology*, vol. 14, p. 102316, 2026.

[14] H. Honda, "Description of the form of trees by the parameters of the tree-like body: Effects of the branching angle and the branch length on the shape of the tree-like body," *Journal of Theoretical Biology*, vol. 31, no. 2, pp. 331–338, 1971.

[15] P. de Reffye, C. Edelin, J. Françon, M. Jaeger, and C. Puech, "Plant models faithful to botanical structure and development," in *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '88)*, 1988, pp. 151–158.

[16] B. Lintermann and O. Deussen, "Interactive modeling of plants," *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 56–65, 1999.

[17] A. Runions, B. Lane, and P. Prusinkiewicz, "Modeling trees with a space colonization algorithm," in *Eurographics Workshop on Natural Phenomena 2007*, 2007, pp. 63–70.

[18] W. Pałubicki, K. Horel, S. Longay, A. Runions, B. Lane, R. Měch, and P. Prusinkiewicz, "Self-organizing tree models for image synthesis," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 58:1–58:10, 2009.

[19] A. Runions, M. Fuhrer, B. Lane, P. Federl, A.-G. Rolland-Lagan, and P. Prusinkiewicz, "Modeling and visualization of leaf venation patterns," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 702–711, 2005.

[20] H. Vogel, "A better way to construct the sunflower head," *Mathematical Biosciences*, vol. 44, no. 3-4, pp. 179–189, 1979.

[21] S. Douady and Y. Couder, "Phyllotaxis as a physical self-organized growth process," *Physical Review Letters*, vol. 68, no. 13, pp. 2098–2101, 1992.

[22] ——, "Phyllotaxis as a dynamical self organizing process. Part I: The spiral modes resulting from time-periodic iterations," *Journal of Theoretical Biology*, vol. 178, no. 3, pp. 255–273, 1996.

[23] O. Šťava, S. Pirk, J. Kratt, B. Chen, R. Měch, O. Deussen, and B. Benes, "Inverse procedural modelling of trees," *Computer Graphics Forum*, vol. 33, no. 6, pp. 118–131, 2014.

[24] A. Raistrick, L. Lipson, Z. Ma, L. Mei, M. Wang, Y. Zuo, K. Kayan, H. Wen, B. Han, Y. Wang, A. Newell, H. Law, A. Goyal, K. Yang, and J. Deng, "Infinite photorealistic worlds using procedural generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[25] B. Gooch and A. Gooch, *Non-Photorealistic Rendering*. Natick, MA: A K Peters, 2001.

[26] T. Strothotte and S. Schlechtweg, *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*. Morgan Kaufmann, 2002.

[27] G. Winkenbach and D. H. Salesin, "Computer-generated pen-and-ink illustration," in *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '94)*, 1994, pp. 91–100.

[28] ——, "Rendering parametric surfaces in pen and ink," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, 1996, pp. 469–476.

[29] M. P. Salisbury, S. E. Anderson, R. Barzel, and D. H. Salesin, "Interactive pen-and-ink illustration," in *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '94)*, 1994, pp. 101–108.

[30] M. P. Salisbury, M. T. Wong, J. F. Hughes, and D. H. Salesin, "Orientable textures for image-based pen-and-ink illustration," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*, 1997, pp. 401–406.

[31] A. Hertzmann and D. Zorin, "Illustrating smooth surfaces," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)*, 2000, pp. 517–526.

[32] O. Deussen and T. Strothotte, "Computer-generated pen-and-ink illustration of trees," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)*, 2000, pp. 13–18.

[33] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. d. O. Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman *et al.*, "Evaluating large language models trained on code," 2021.

[34] B. Zou, M. Cai, J. Zhang, and Y. J. Lee, "VGBench: Evaluating large language models on vector graphics understanding and generation," 2024.

[35] R. Wu, W. Su, and J. Liao, "Chat2SVG: Vector graphics generation with large language models and image diffusion models," 2024.

[36] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prabhumoye, Y. Yang, S. Gupta, B. P. Majumder, K. Hermann, S. Welleck, A. Yazdanbakhsh, and P. Clark, "Self-refine: Iterative refinement with self-feedback," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[37] N. Shinn, F. Cassano, E. Berman, A. Gopinath, K. Narasimhan, and S. Yao, "Reflexion: Language agents with verbal reinforcement learning," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[38] K. Ellis, D. Ritchie, A. Solar-Lezama, and J. B. Tenenbaum, "Learning to infer graphics programs from hand-drawn images," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[39] W. Blunt and W. T. Stearn, *The Art of Botanical Illustration*. Woodbridge: Antique Collectors' Club, in

association with the Royal Botanic Gardens, Kew, 1994, new edition.

[40] G. Saunders, *Picturing Plants: An Analytical History of Botanical Illustration*. Berkeley: University of California Press, in association with the Victoria and Albert Museum, 1995.

[41] F. E. Köhler, *Köhler's Medizinal-Pflanzen in naturgetreuen Abbildungen mit kurz erläuterndem Texte*, G. Pabst, Ed. Gera-Untermhaus: Franz Eugen Köhler, 1887, digitized by the Biodiversity Heritage Library, https://www.biodiversitylibrary.org/bibliography/623.

[42] "Curtis's botanical magazine," 1787, continuously published since 1787 by the Royal Botanic Gardens, Kew, currently published by Wiley. Print ISSN 1355-4905, online ISSN 1467-8748.

[43] E. R. S. Hodges, Ed., *The Guild Handbook of Scientific Illustration*, 2nd ed. John Wiley & Sons, 2003.

[44] T. Isenberg, P. Neumann, S. Carpendale, M. C. Sousa, and J. A. Jorge, "Non-photorealistic rendering in context: an observational study," in *Proceedings of the 4th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '06)*, 2006, pp. 115–126.

[45] A. Hertzmann, "Non-photorealistic rendering and the science of art," in *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering (NPAR '10)*, 2010, pp. 147–157.

[46] T. Ettinger, "Mulberry32: A tiny, fast, 32-bit pseudo-random number generator," https://gist.github.com/tommyettinger/46a874533244883189143505d203312c, 2017, gitHub Gist, accessed 2026-09-23.

[47] A. Jain, A. Xie, and P. Abbeel, "VectorFusion: Text-to-SVG by abstracting pixel-based diffusion models," 2022.