

An open knowledge base for

INDIC TEXT SHAPING

Bonus: A trip through a wormhole.

梁海 Liang Hai, independent researcher: lianghai.github.io

This ongoing project is sponsored by Typotheque: github.com/typotheque/text-shaping

IUC 44, October 2020



- analytical framework
- production tutorials
- reference font sources
- glyph database
- configurable tooling
- community participation
- standardization

Indic Text Shaping for Type Designers

github.com/typotheque/text-shaping

ARCHITECTURAL CONTEXT

analog text

display

input

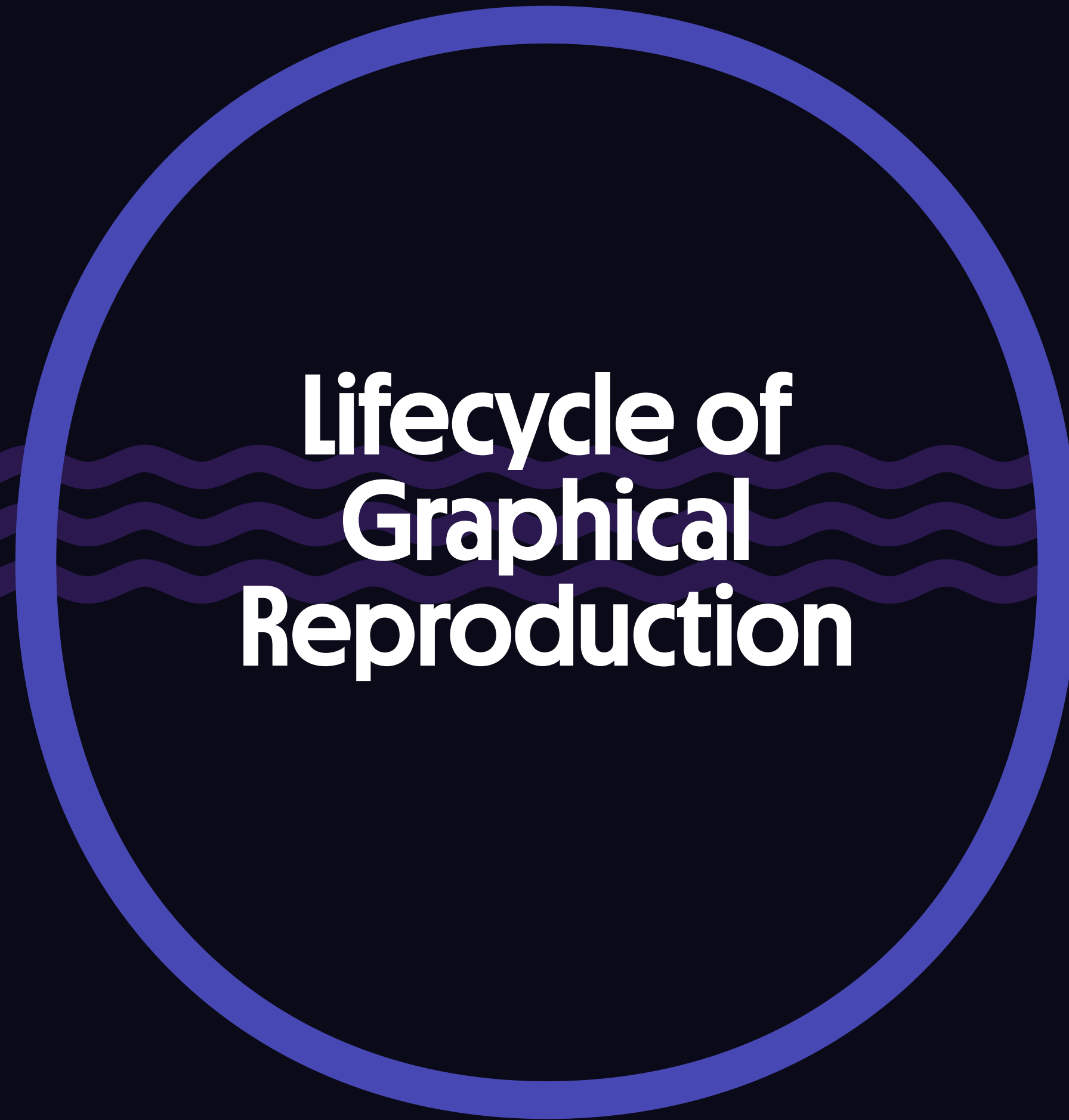


digital text

analog text

Tangible

display



input

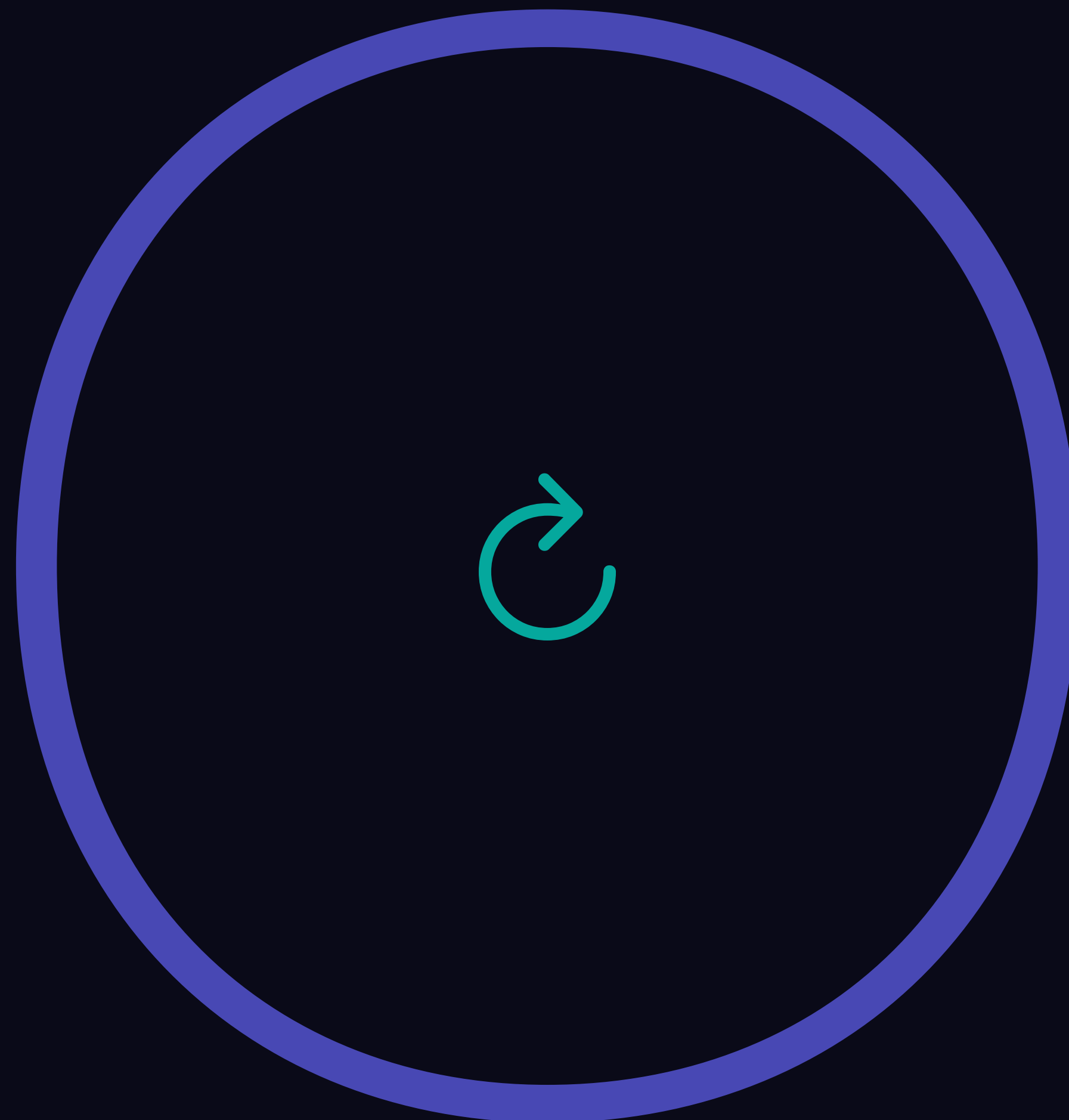


digital text

Abstract

text as images

shaping with
OpenType Layout
technology



encoding per the
Unicode Standard's
recommendation

text representation
in *Unicode* characters

« touché »

U+0074 → \t
U+006F → \o
U+0075 → \u
U+0063 → \c
U+0068 → \h
U+00E9 → \eacute



U+0074 t LATIN SMALL LETTER T
U+006F o ... LETTER O
U+0075 u ... LETTER U
U+0063 c ... LETTER C
U+0068 h ... LETTER H
U+00E9 é ... LETTER E WITH ACUTE

<0074, 006F, 0075, 0063, 0068, 00E9>

« touché »

U+0074 → |t
U+006F → |o
U+0075 → |u
U+0063 → |c
U+0068 → |h
U+0065 → |e
U+0301 → |acutecomb



U+0074 t LATIN SMALL LETTER T
U+006F o ... LETTER O
U+0075 u ... LETTER U
U+0063 c ... LETTER C
U+0068 h ... LETTER H
U+0065 e ... LETTER E
U+0301 ó COMBINING ACUTE ACCENT

<0074, 006F, 0075, 0063, 0068, 0065, 0301>

« कॉम्प्लेक्स लिपि »

- reordering
- conjoining
- contextual variation
- productivity
- prioritization
- special cases
- edge case handling
- locale-specific variation
- ...



- graphical segmentation
- *akshara*-based phonetic aid
- double-acting vowel killer
- ZWJ/ZWNJ override
- ...

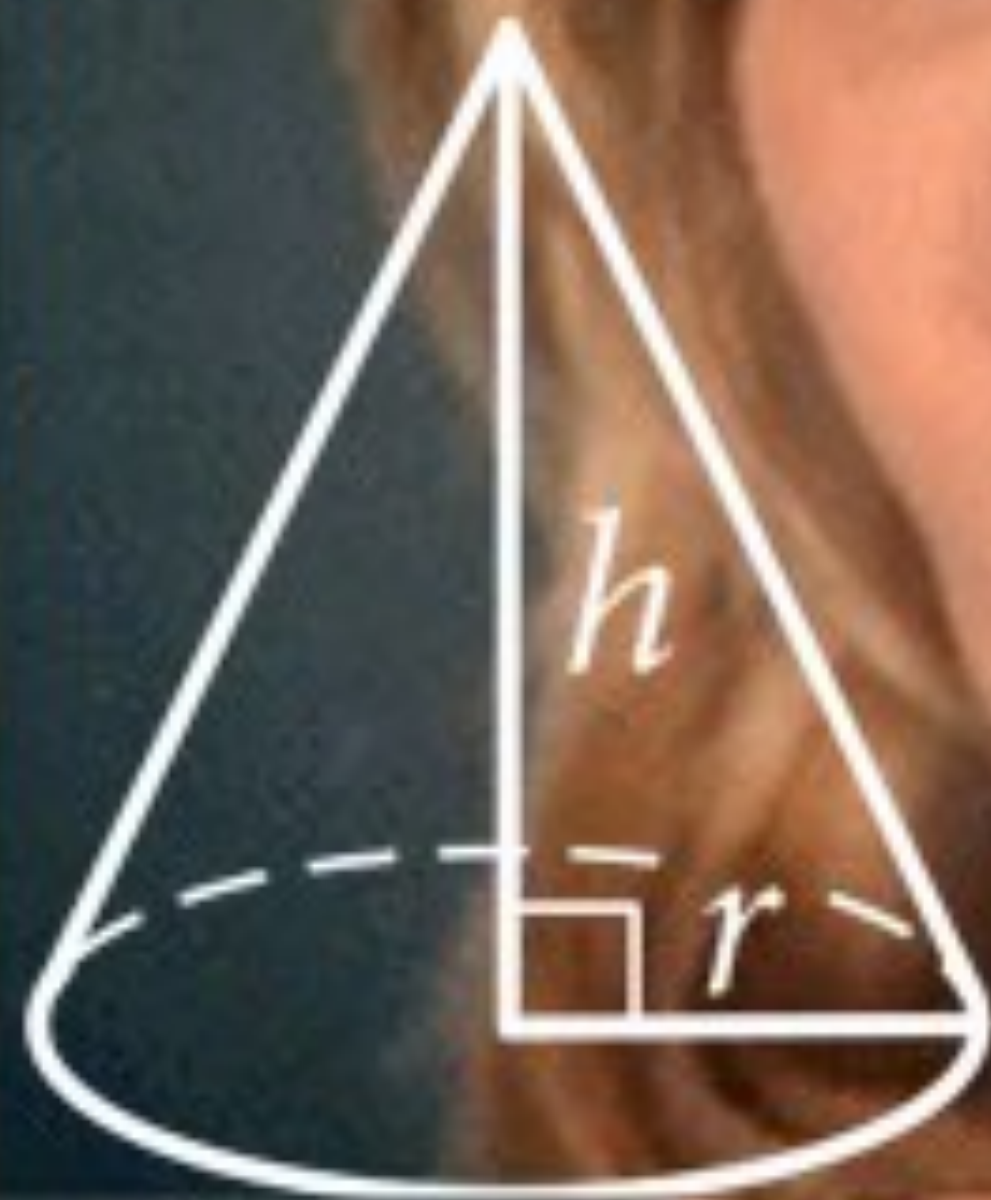
<0915, 0949, 092E, 094D, 092A, 094D, 0932, 0947, 0915, 094D, 0938, ...>



$$A = \pi r^2$$

$$C = 2\pi r$$

$$V = \frac{1}{3} \pi r^2 h$$



$$V = \pi r^2 h$$

	30°	45°	60°
sin	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$
cos	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$
tan	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$



$$\int \sin x dx = -\cos x + C$$

$$\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C$$

$$\int \operatorname{tg} x dx = -\ln|\cos x| + C$$

$$\int \frac{dx}{\sin x} = \ln\left|\operatorname{tg} \frac{x}{2}\right| + C$$



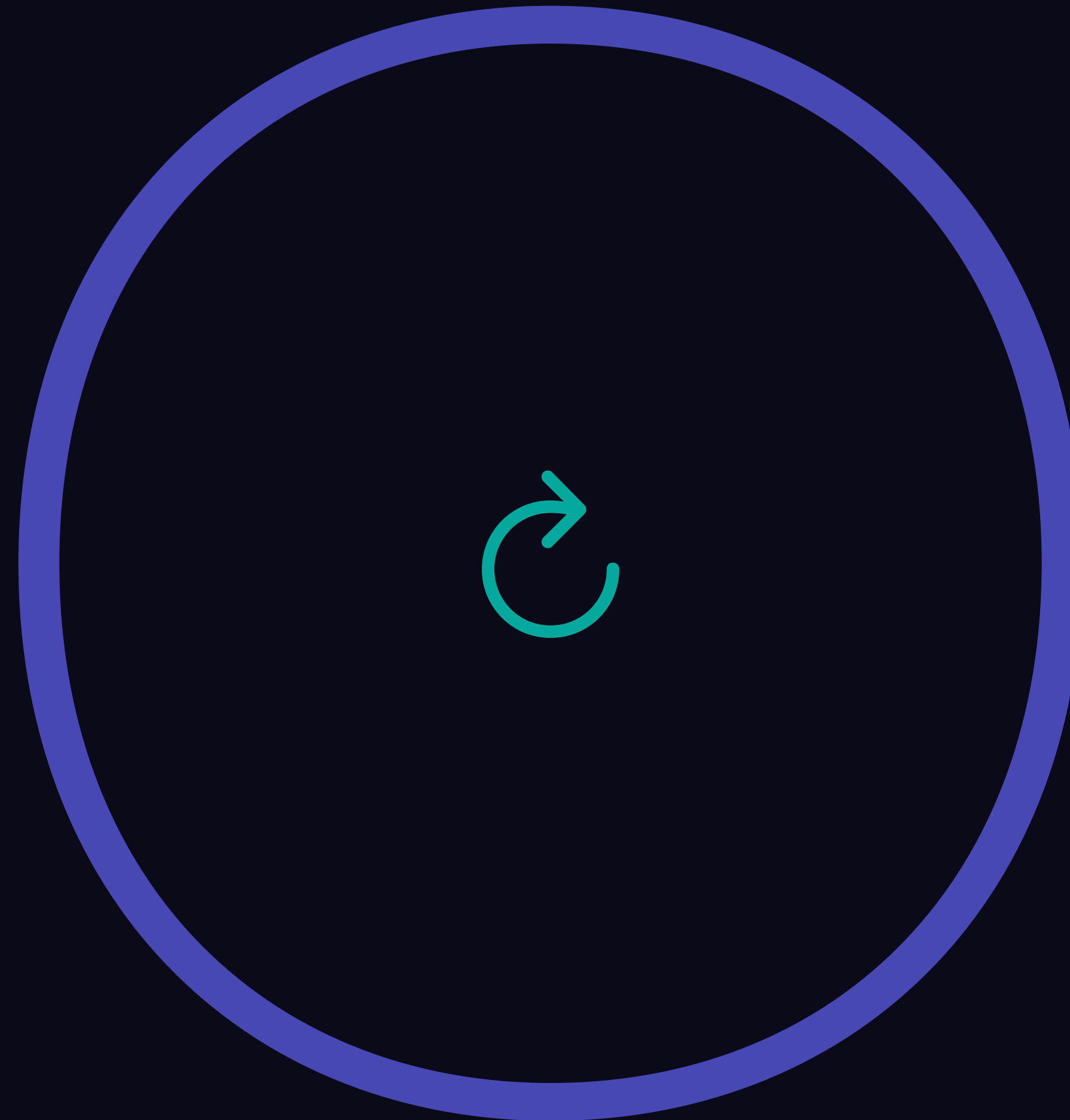
$$ax^2 + bx + c = 0$$

$$a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) = 0$$

ABSTRACTION LAYER

text as images

shaping with
OpenType Layout
technology



encoding per the
Unicode Standard's
recommendation

text representation
in Unicode characters

text as images



typographical
& dynamic

shaping

orthographical
& critical

graphical
analysis

encoding

model-specific
analysis

text representation
in Unicode characters

text as images



typographical
& dynamic

shaping

orthographical
& critical

graphical
analysis

encoding

model-specific
analysis

text representation
in Unicode characters

text as images

*late shaping:
typographical
& dynamic*

*early encoding:
graphical
analysis*

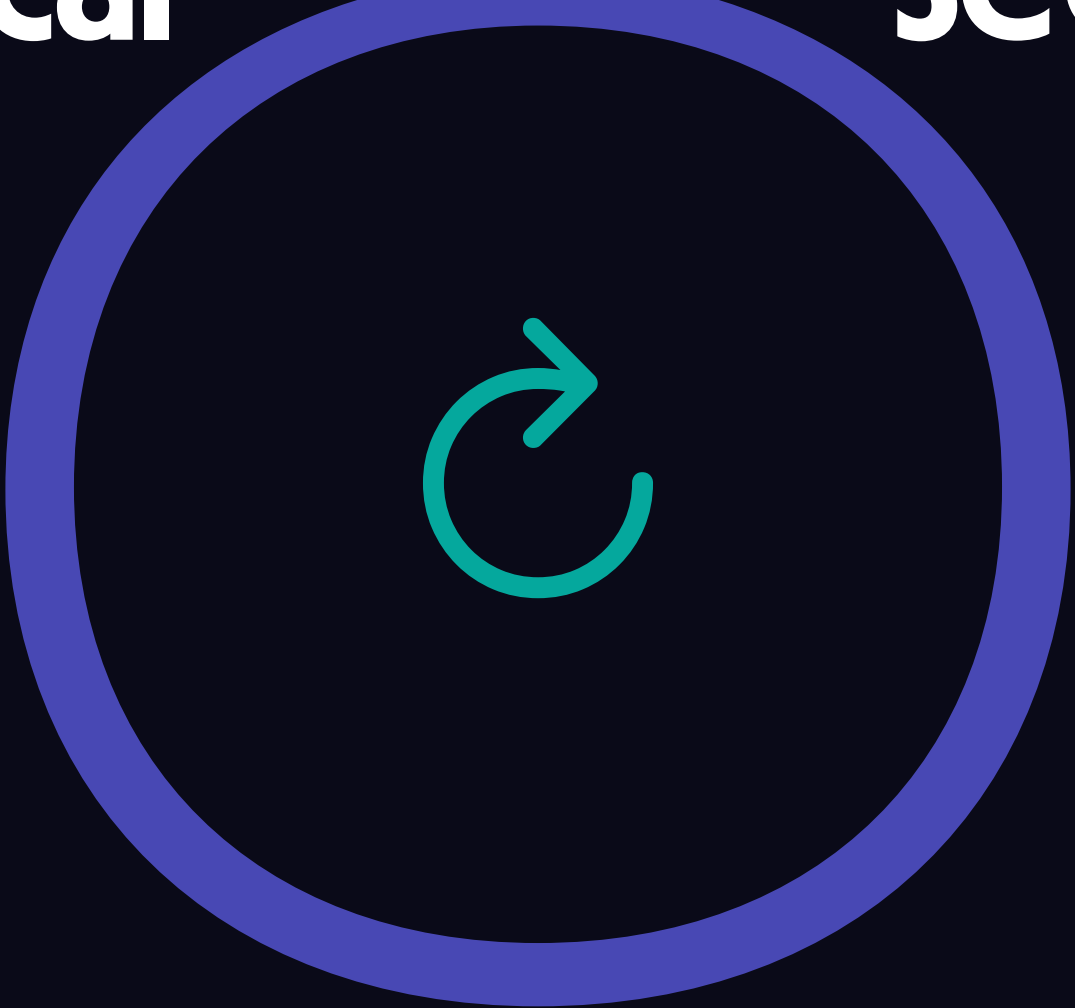
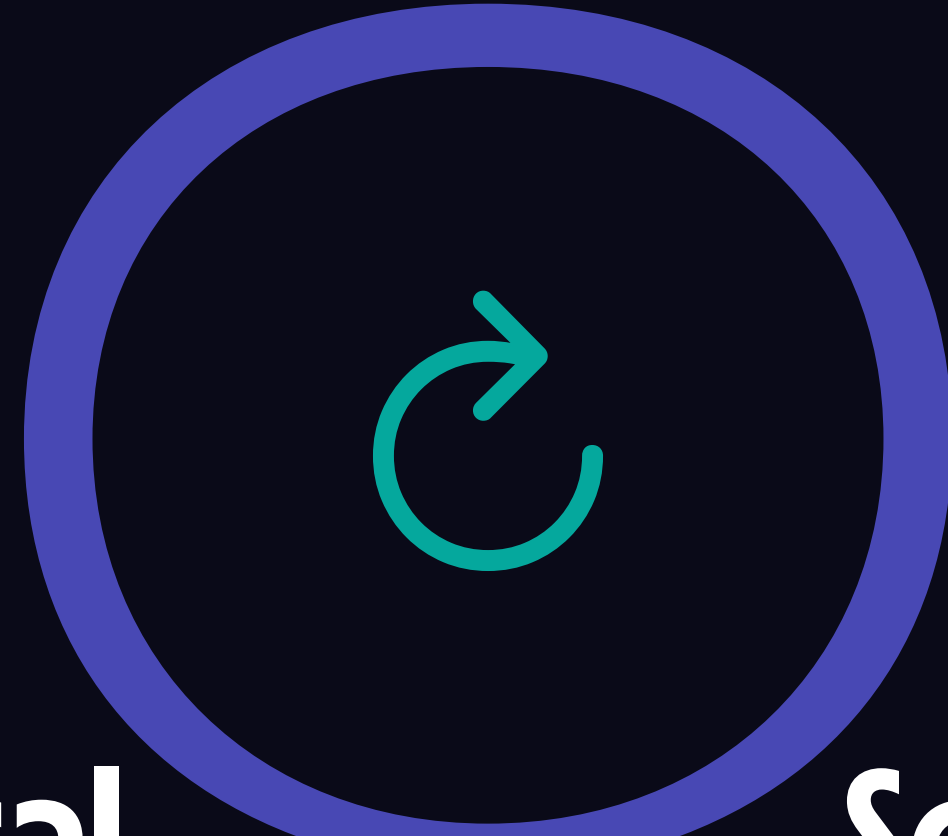
Graphical

Segmentation

*early shaping:
orthographical
& critical*

*late encoding:
model-specific
analyses*

text representation
in Unicode characters



कॉम्प्लेक्स लिपि

text as images

**Graphical
Segmentation**



text representation
in Unicode characters



कॉम्प्लेक्स लिपि

text as images

Graphical
Segmentation



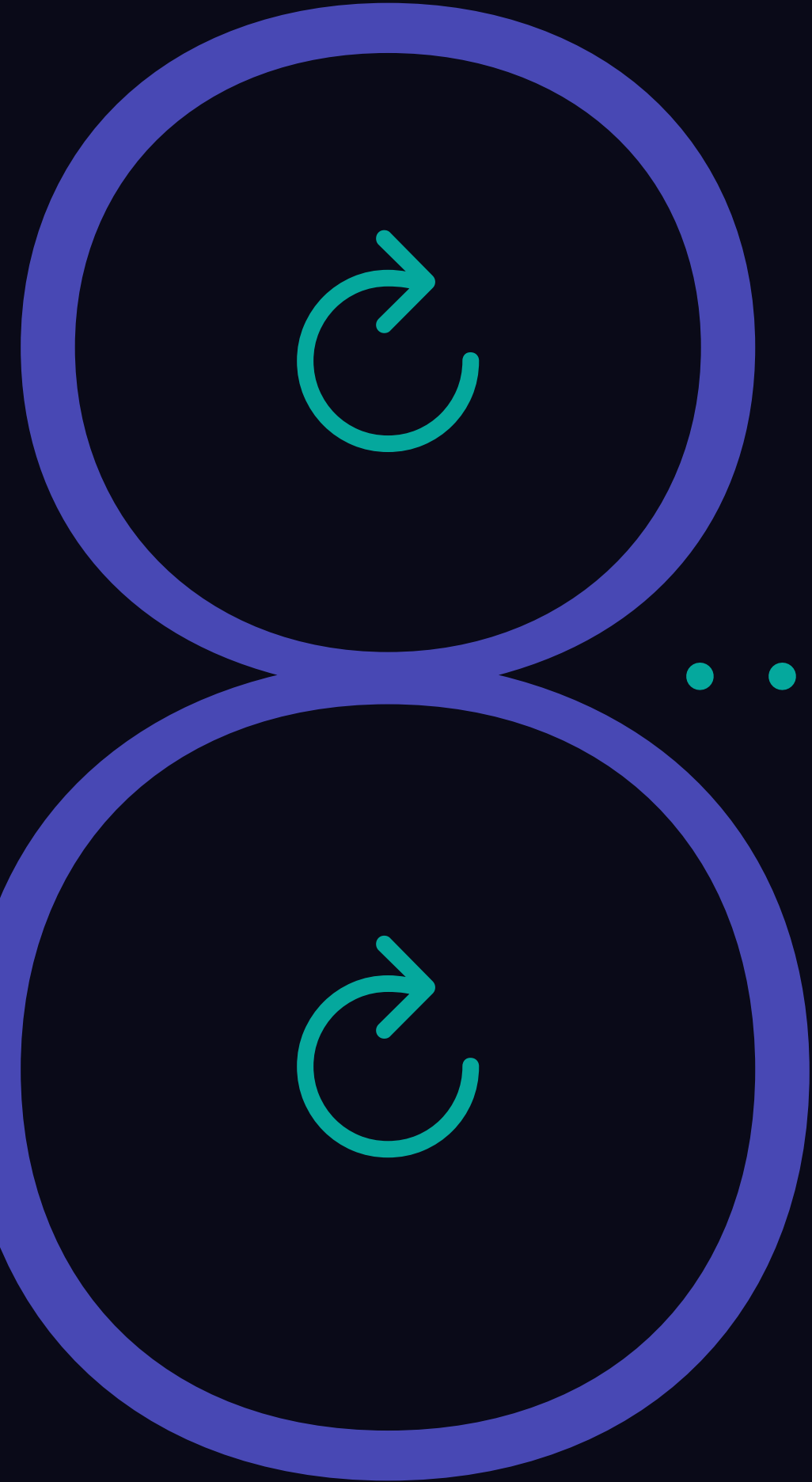
text representation
in Unicode characters

कॉम्प्लेक्स
लिपि

कॉम्प्लेक्स लिपि

text as images

Graphical
Segmentation



कॉम्प्लेक्स लिपि

text representation
in Unicode characters

कॉम्प्लेक्स
लिपि

**DON'T BE AFRAID
TO LOOK BACK**

text as images

comprehensible &
reusable foundation

maximal
textual abstraction

Graphical

Segmentation

relatively stable &
optimizable target

reliable
reference point

text representation
in Unicode characters



कॉम्प्लेक्स लिपि User Space

Any text processes that rely on graphical analysis.

- type design
- layout engines
- input method UI
- ...

Graphical
Segmentation



..... कॉम्प्लेक्स लिपि

Graphical
Segmentation



कॉम्प्लेक्स लिपि

The industry needs to own its encoding mess and provide solutions.

- systematic glyph names
- reference glyph sets
- reference shaping rules
- shaping engines
- encoding guidelines

...

Infrastructure Space

कॉम्प्लेक्स
लिपि



User Space

Graphical Segmentation



Infrastructure Space

Indic Text Shaping for Type Designers

github.com/typotheque/text-shaping