# pdftomd.sh Activity Diagram

```
                              ●
                              │
                              ▼
                      ┌───────────────┐
                      │  Load defaults │
                      └───────────────┘
                              │
                              ▼
                ┌───────────────────────────┐
                │ Load pdftomd.conf if present│
                └───────────────────────────┘
                              │
                              ▼
                      ┌───────────────┐
                      │ Parse CLI args │
                      └───────────────┘
                              │
                              ▼
                  ┌─────────────────────┐
                  │ Validate required args│
                  └─────────────────────┘
                              │
                              ▼
                    ┌───────────────────┐
                    │ Print flag summary │
                    └───────────────────┘
                              │
        yes                   ◇ Input is directory?        no
   ┌──────────────────────────┴──────────────────────────┐
   ▼                                                      ▼
┌────────────────────────┐                      ┌──────────────────────┐
│ Announce directory mode │                      │  Resolve input path   │
└────────────────────────┘                      └──────────────────────┘
   │                                                      │
   ▼                                                      ▼
┌────────────────────────────┐                   ◇ Input file exists?   yes
│ Collect PDFs (recursive if -r)│                  │ no
└────────────────────────────┘                    ▼
   │                                          ┌────────┐
   ▼                                          │ Exit 1 │
┌──────────────────────────────────────┐     └────────┘
│ For each PDF -> re-run script with     │        │
│ same flags                             │        ▼
└──────────────────────────────────────┘        ◉
   │
   ◇ Child exit code >= 128?
   │ yes
   ▼
┌──────────────────────────┐
│ Stop directory processing │
└──────────────────────────┘
   │
   ▼
   ◉
   │
   ▼
┌──────────────┐
│ Track failures│
└──────────────┘
   │
   ◇ Any failures?   no   ◉
   │ yes
   ▼
┌────────┐
│ Exit 1 │
└────────┘
   │
   ▼
   ◉
                              ┌──────────────────────┐
                              │  Init timers/logging  │
                              └──────────────────────┘
```

```
              ┌─────────────┐
              │  -o/--ocr?  │────────────┐
              └─────────────┘            │
                  │ yes                  │
           ┌────────────────┐            │
           │ Run ocr-pdf.sh │            │
           └────────────────┘            │
                  │                      │
         ┌────────────────────┐          │
         │ OCR output missing?│──────┐   │
         └────────────────────┘      │   │
                  │ yes              │   │
              ┌────────┐            │   │
              │ Exit 1 │            │   │
              └────────┘            │   │
                  │                 │   │
                  ●                 │   │
                                    │   │
              ┌───────────────────────┐
              │ Use _OCR.pdf as source│
              └───────────────────────┘
                       │
                      ◇──────────────────┘
                       │
         ┌──────────────────────────┐
         │ strip mode forced AND -o?│──────┐
         └──────────────────────────┘      │
                  │ yes                    │
           ┌──────────────────┐            │
           │ Disable strip mode│           │
           └──────────────────┘            │
                  │                        │
                 ◇─────────────────────────┘
                  │
         ┌─────────────┐
         │ -o not used │──────────────────────────────────┐
         └─────────────┘                                  │
                  │ yes                                    │
    ┌──────────────────┐        ┌──────────────────┐      │
    │ strip mode = force?│─────►│ strip mode = auto?│─────┐│
    └──────────────────┘        └──────────────────┘     ││
           │ yes                       │ yes             ││
  ┌──────────────────┐        ┌────────────────┐        ││
  │Strip OCR layer   │        │ Detect OCR layer│        ││
  │(fast)            │        └────────────────┘        ││
  └──────────────────┘               │                  ││
           │                   ┌───────────┐             ││
  ┌──────────────────────┐    │ Detected? │──────┐       ││
  │Warnings -> malformed │    └───────────┘      │       ││
  │PDF?                  │──┐       │ yes        │       ││
  └──────────────────────┘  │  ┌────────────────┐       ││
           │ yes           │  │Strip OCR layer  │       ││
  ┌──────────────────┐     │  │(fast)           │       ││
  │Attempt qpdf      │     │  └────────────────┘       ││
  │--repair          │     │       │                    ││
  └──────────────────┘     │  ┌──────────────────────┐  ││
           │               │  │Warnings -> malformed │  ││
  ┌──────────────────┐     │  │PDF?                  │──┐││
  │ Retry strip once │     │  └──────────────────────┘  │││
  └──────────────────┘     │       │ yes                │││
           │               │  ┌──────────────────┐      │││
          ◇────────────────┘  │Attempt qpdf      │      │││
           │                  │--repair          │      │││
                              └──────────────────┘      │││
                                   │                     │││
                              ┌──────────────────┐       │││
                              │ Retry strip once │       │││
                              └──────────────────┘       │││
```

Split PDF into chunks

Prepare Marker args

-l/--llm?

yes

Enable LLM helper

-o/--ocr?

yes

no

Pass --disable_ocr to Marker

Write config_json {force_ocr=true, strip_existing_ocr=depends on strip mode}

Run Marker on chunk dir

Rate limit error and -l?

yes

Retry once without --use_llm

Marker failed?

yes

Report error and exit 1

Merge chunk markdown

Restore images/attachments

Move output to start directory

--preclean-copy? — yes

Save _preclean.md copy

--clean? — yes

Run LLM cleanup

Cleanup fails? — yes

Restore pre-clean markdown

Print timing stats